DOCUMENT RESUME

ED 436 723 CS 013 773

AUTHOR Taylor, Barbara M.; Pearson, P. David; Clark, Kathleen F.; Walpole, Sharon

TITLE Beating the Odds in Teaching All Children To Read.

INSTITUTION Center for the Improvement of Early Reading Achievement, Ann Arbor, MI.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO CIERA-R-2-006
PUB DATE 1999-09-30
NOTE 64p.
CONTRACT R305R70004
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Early Intervention; Independent Reading; *Parent School Relationship; Phonics; Primary Education; *Reading Achievement; *School Effectiveness; *Teacher Behavior; *Teacher Effectiveness; Word Recognition
IDENTIFIERS *Effective Practice Identification; *Effective Reading Programs

ABSTRACT

A study investigated school and classroom factors related to primary-grade reading achievement, using quantitative and descriptive methods. Fourteen schools across the United States with moderate to high numbers of students on subsidized lunch were identified as most, moderately, or least effective based on several measures of reading achievement. A combination of school and teacher factors was found to be important in the most effective schools. Significant factors included: (1) strong links to parents; (2) systematic assessment of pupil progress; (3) strong building communication; (4) a collaborative model for reading instruction, including early reading interventions; (5) time spent in small group instruction; (6) time spent in independent reading; (7) high pupil engagement; and (8) strong home communication. The most accomplished teachers were frequently observed teaching word recognition by coaching as children were reading, providing explicit phonics instruction, and asking higher level questions after reading. In all of the most effective schools, reading was clearly a priority at both the building and classroom levels. Contains 49 references and 21 tables. Appendixes provide sample worksheets and instructions for administering word lists. (EF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 436 723

Beating the Odds in Teaching All Children to Read

Barbara M. Taylor, University of Minnesota
P. David Pearson, Michigan State University
Kathleen F. Clark, University of Minnesota
Sharon Walpole, University of Virginia

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CIERA REPORT #2-006
September 30, 1999

CIERA
Center for the Improvement of Early Reading Achievement

CS 013 773

Beating the Odds in Teaching All Children to Read

CIERA REPORT #2-006

Barbara M. Taylor, University of Minnesota
P. David Pearson, Michigan State University
Kathleen F. Clark, University of Minnesota
Sharon Walpole, University of Virginia

CIERA Inquiry 2: Home and School
What schoolwide practices characterize schools in which at-risk learners are beating the odds? What instructional practices are used by the most accomplished primary-grade teachers and by teachers in the most effective schools?

The authors used quantitative and descriptive methods to investigate school and classroom factors related to primary-grade reading achievement. Fourteen schools across the U.S. with moderate to high numbers of students on subsidized lunch were identified as most, moderately, or least effective based on several measures of reading achievement in the primary grades. A combination of school and teacher factors, many of which were intertwined, was found to be important in the most effective schools. Statistically significant school factors included strong links to parents, systematic assessment of pupil progress, strong building communication, and a collaborative model for the delivery of reading instruction, including early reading interventions. Statistically significant teacher factors included time spent in small group instruction, time spent in independent reading, high pupil engagement, and strong home communication. More of the most accomplished teachers were frequently observed teaching word recognition by coaching as children were reading, in addition to providing explicit phonics instruction, than the least accomplished teachers and teachers in the moderately or least effective schools. Additionally, more of the most accomplished teachers and those in effective schools were frequently observed asking higher level questions after reading than their counterparts. In all of the most effective schools, reading was clearly a priority at both the building and classroom level.

Center for the Improvement of Early Reading Achievement
University of Michigan – Ann Arbor

September 30, 1999

CIERA
Center for the Improvement of Early Reading Achievement

University of Michigan School of Education

610 E University Ave., Rm 1600 SEB
Ann Arbor, MI 48109-1259

734.647.6940 voice
734.763.1229 fax
ciera@umich.edu

www.ciera.org

©1999 Center for the Improvement of Early Reading Achievement.

This research was supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the comments do not necessarily represent the positions or policies of the National Institute of Student Achievement, Curriculum, and Assessment or the National Institute on Early Childhood Development, or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

Beating the Odds in Teaching All Children to Read

Barbara M. Taylor, University of Minnesota
P. David Pearson, Michigan State University
Kathleen F. Clark, University of Minnesota
Sharon Walpole, University of Virginia

We all want the best schools possible for our children, schools that help them acquire the knowledge, skills, and dispositions they will need to pursue whatever dreams and paths they wish. Yet the reality is that many of our children are not reading well enough to keep up with the demands of school (Campbell, Donahue, Reese, & Phillips, 1996; Donahue, Voelkl, Campbell, & Mazzeo, 1999), let alone the demands of our society or their personal dreams. In the recent national report, *Preventing Reading Difficulties in Young Children*, a National Academy of Science Committee concluded that "quality classroom instruction in kindergarten and the primary grades is the single best weapon against reading failure" (Snow, Burns, & Griffin, 1998). The committee recommended that our number one priority for funding research should be to improve classroom reading instruction in kindergarten and the primary grades.

In a recent report of a three-year study of schools implementing special strategies to improve reading achievement, researchers described classroom instruction as "in one sense, distressing" (Stringfield, Millsap, & Herman, 1997, p. 2). In the elementary schools, instruction was predominantly teacher-led, focused on discrete skill instruction, and driven by management concerns. There were relatively few observations of students engaged in sustained reading or students applying what they were learning. On the other hand, Stringfield, Millsap, and Herman pointed out that even in schools nominated as exemplary, there was ample room for instructional improvement, which would, if implemented, lead to greater gains in reading achievement.

In addition to advocating improved classroom reading instruction, the Committee on the Prevention of Reading Difficulties in Young Children discussed the importance of systematic, schoolwide restructuring efforts in reading. The committee (Snow et al., 1998) recommended that poor performing schools consider reading reform efforts with a dual focus on schoolwide organizational issues *and* improved classroom reading instruction.

In their special strategies study, Stringfield et al. (1997) found that reform programs that focused on the primary grades had larger achievement gains

in reading than schools that spread their efforts out across the elementary grades or into the secondary grades. They also found that schools that adopted externally developed programs had greater achievement gains than schools that developed their own programs.

In yet another recent report of a large national study of 400 Chapter 1 schools, researchers found that higher levels of poverty, greater application of grade retention policies, and higher levels of student disciplinary actions were related to lower student achievement (Puma et al., 1997). Only 5 schools in the pool of 400 were identified as exceptional. These schools tended to have a "more experienced principal, a schoolwide Chapter 1 program, some tracking by ability in grades 1-6, lower rates of teacher and student mobility, a balanced emphasis on remedial and higher-order thinking in classroom instruction, and higher levels of community and parent support" (p. 62). Except for first grade, in which grouping was used, whole-class instruction was the dominant practice in all schools.

These three recent national reports highlight the importance of and need for additional research on schools that serve the needs of poor children by increasing their achievement and, hence, their educational opportunities. The purpose of the present study was to examine the instructional and organizational factors that might explain how and why some schools across the country are *beating the odds* by attaining greater than expected primary-grade reading achievement with populations of students at risk for failure by virtue of poverty. We pause to emphasize the terms *instructional* and *organizational*, for it is our belief that only when we attend to both school-level (organizational) and classroom-level (instructional) facets of reform do we meet our aspirations.

Within this broader framework, we were, like the researchers in the special strategies study (Stringfield et al., 1997), interested in both *imported* models of reform (where they had adopted an external intervention program or school reform program) and *homegrown* reform efforts. To that end, we sought schools in both categories. Among the reforms imported were schoolwide programs and both tutorial and small-group interventions for at-risk youngsters. As it turned out, we also had *combination* reform efforts; in several schools, for example, an imported intervention program was set within a homegrown schoolwide reform effort.

Effective Schools

Research on effective schools relevant to reading achievement, much of which was conducted in the 1970s and early 1980s, was documented in a review entitled "Teacher and School Effects in Learning to Read" by Hoffman (1991) in the *Handbook of Reading Research, Volume II*. Hoffman described eight attributes of effective schools frequently summarized in the literature (e.g., Shavelson & Berliner, 1988), including:

- A clear school mission;
- effective instructional leadership and practices;

- high expectations;
- a safe, orderly, and positive environment;
- ongoing curriculum improvement;
- maximum use of instructional time;
- frequent monitoring of student progress; and
- positive home-school relationships.

In a study of 4 outlier inner-city schools, Weber (1971) found strong leadership, high expectations, positive school climate, strong emphasis on reading, and continuous evaluation of pupil progress related to the identified school reading success criteria of median achievement level on a normed standardized reading achievement test and having a relatively small number of children with serious reading difficulties. In a study of 5 schools found to be most effective out of a sample of 741 schools which were part of a study of compensatory reading programs, Wilder (1977) found the following factors common to all 5 schools: reading was identified as an important instructional goal; leadership in the reading program was provided by either the principal or reading specialist; attention was given to basic skills; a breadth of materials was made available; and ideas were communicated across teachers, a process which was typically fostered by the program leader.

In a more recent, longitudinal study on schools implementing special strategies for educating disadvantaged children, Stringfield et al. (1997) found that the schools demonstrating the greatest achievement gains worked hard at both initial implementation and long-term maintenance of an innovation. But the researchers also noted the importance of systematic self-improvement in these schools, in which the innovations continued to evolve and expand. Externally developed research-based programs and programs that focused on whole school reform were related to greater achievement gains than locally developed programs and innovations composed of various pull-out programs. The study also found support for the premise that students placed at risk of academic failure could achieve at levels that met national averages.

In a study of five effective Title I schools, Puma et al. (1997) found that high-performing, high-poverty schools had lower than average teacher and student mobility, principals with more years of experience, and more orderly school environments than average high-poverty schools. Better school climates and better relations with administration and the community were also reported, as well as greater parent involvement and more parents with high expectations for their children's future educational attainment. All of these high-performing schools had tracking by ability in grades 1–6. In three of the five schools, teachers emphasized remedial and higher order comprehension skills in reading.

Although research on effective schools has been favorably received by school leaders and policymakers, Hoffman (1991) points to limitations of this research stemming from its lack of connection to classroom practice and to insufficient information on the process schools went through to become effective. Even so, that the same characteristics arise time and time

again has led many reformers to suggest that these findings ought to be translated into policies that can guide reform.

Effective Teachers

In addition to research on effective schools, Hoffman (1991) summarized a considerable body of research, spanning the 1960s through the 1980s, on teachers who were exceptionally effective in helping students learn to read. Hoffman reports on a literature review of effective teaching by Rosenshine and Furst (1973) in which they found several teacher behaviors consistently related to student achievement: clarity, variability, enthusiasm, task orientation, teacher directness, student opportunity to learn criterion material, use of structuring comments, multiple levels of questions, and criticism (which was negatively related to achievement).

In a study of the achievement of students of 165 second- and third-grade teachers conducted over a three-year period, Brophy (1973) reported on the patterns of the most effective teachers, who represented about a third of the sample. He found that the most effective teachers were businesslike with a strong sense of task and direction for themselves and their students, had high expectations for their students' achievement, and redoubled efforts when failure was experienced, especially in low socioeconomic status (SES) environments. The most effective teachers had strong management skills, but their classrooms were not stern or oppressive. They had high levels of pupil engagement and were proactive in preventing disruptions. The most effective teachers engaged in the practice of probing individuals when incorrect responses were offered instead of simply calling on someone else or giving the answer themselves. The students in low-SES classes of the most effective teachers had a success rate of about 80% correct when answering teacher-directed questions, almost all of which were literal. In a follow-up intervention study of first-grade teachers engaged in small-group instruction, Anderson, Evertson, and Brophy (1979) found that greater achievement was related to more time spent in reading groups, more active instruction, shorter transitions, introduction of lessons with an overview, and follow-up by teachers to incorrect responses with attempts to improve upon them.

In a study of 166 first- and third-grade teachers of children who had been in Head Start, variables positively related to gains in reading included time spent in academic activities, frequency of small-group instruction in basic skills, and frequency of supervised seatwork activities (Stallings & Kaskowitz, 1974). The lowest SES students benefited most from intense, small-group instruction.

In a study of 25 second-grade and 21 fifth-grade classrooms, Fisher et al. (1980) found that the more effective teachers had higher amounts of time allocated to academics and higher pupil engagement than less effective teachers. High success rates on tasks were also found to be related to learning gains, with higher optimum success rates found for low-ability than for high-ability students.

Knapp (1995) studied 140 grade 1-6 classrooms in 15 high-poverty schools in California, Maryland, and Ohio over a two-year period. They found that students in grades 1, 3, and 5 who were exposed to meaning-oriented reading instruction performed 5.6 national curve equivalents (NCEs) higher, and students in grades 2, 4, and 6, 1.4 NCEs higher, at the end of the school year than students in classrooms with skills-oriented approaches to reading instruction. They also studied effects in math and writing and concluded that meaning-oriented instruction was effective in high-poverty classrooms. The teachers they observed teaching for meaning wanted to give children more responsibility for learning, wanted to provide academic tasks that asked more of students, and sustained engagement in learning among children.

The work of Wharton-MacDonald, Pressley, and Hampston (1998) both echoes and extends the earlier research on effective teachers of beginning reading. Three of the nine first-grade urban teachers in their sample were identified as most effective based on their students' end-of-year reading and writing achievement. These teachers demonstrated instructional balance, focusing on both literature and skills. They taught decoding skills explicitly and also provided their students with many opportunities to engage in authentic, integrated reading and writing activities. In contrast, the other teachers in the study either focused on skills or whole language approaches or combined the two in disjointed ways. The three most effective teachers extensively used scaffolding to help their students learn. They encouraged self-regulation by teaching their students to monitor their learning, the quality of their work, and their work time. They also encouraged self-regulation by teaching students to use strategies to be good readers and to fix problems they encountered as they were reading. The best teachers had high expectations for their students and masterful classroom management skills. They were skilled in managing time as well as behavior. They were well prepared for their lessons, and they mentioned the importance of routines in terms of activities and expectations. Finally, the most effective teachers were clear about the purposes of their activities and practices.

In the conclusion to their paper, Wharton-MacDonald et al. (1998) pointed to the need for additional research on the role of school factors and district policies on teacher practices and student performance. Hoffman (1991) also observed that there has been a paucity of research simultaneously investigating *both* school and classroom factors affecting reading achievement. Clearly, more research operating at the effective school–effective teacher nexus is needed. Such research would, in a single effort, examine school level factors (e.g., building climate, home-school relations, schoolwide organization for reading, collaborative efforts) while examining classroom/teacher factors (e.g., time spent in reading instruction, time on task, student engagement, approaches to word recognition and comprehension instruction, teachers' interactive styles).

In this study, we attempted to wed these important, but seldom integrated, lines of inquiry. We used quantitative and descriptive methods to examine the programs and practices in 11 moderate- to high-poverty schools selected because of their dual reputation for implementing recent reading reform and for beating the odds by promoting greater than expected primary-grade reading achievement. We also examined three schools chosen because they allegedly produced rather ordinary achievement. However, during the

course of the data collection, some schools surfaced as more effective than others (see Stringfield et al., 1997, for a similar phenomenon). Therefore, rather than rely on a priori labels, we sought to pinpoint and explain school level (i.e., program) and classroom level (i.e., teachers' instructional practices) factors that distinguished the most effective schools from other schools in the study.

Method

Participants

Fourteen schools geographically dispersed throughout the country took part in the study, including schools in Virginia, Minnesota, Colorado, and California. A summary of the characteristics of each school, including type of intervention and type of schoolwide innovation, if any, in reading appears in Table 1. Schools in the study ranged from 28–92% poverty, and included four rural, four small town, and one suburban school, as well as five inner-city schools from three large metropolitan districts.

We started by trying to identify schools with two characteristics: (a) those that had recently implemented reform programs to improve reading achievement, and (b) those with a reputation for producing unexpectedly positive results with low-income populations. Because we were interested in special interventions for students most at risk for failure, we selected 8 schools which had carefully implemented an externally developed, research-proven early reading intervention, including 1 Book Buddies school (Invernizzi, Juel, & Rosemary, 1997), 2 Early Intervention in Reading schools (Taylor, Short, Frye, & Shearer, 1992), 3 schools with Right Start in Reading (Hiebert, Colt, Catto, & Gury, 1992), and 2 Reading Recovery schools (Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994). In 6 of these schools, the interventions were set within the broader context of program reform. The other 2 schools in this group of 8 had implemented early reading interventions without schoolwide reform of their reading program. Additionally, we selected two schools which had implemented externally developed, nationally recognized schoolwide reform programs—Success for All (Madden, Slavin, Karweit, Dolan, & Wasik, 1993) and Core Knowledge (Hirsch, 1987). Also included in the sample was one homegrown reform school; it had neither an externally developed schoolwide reform nor an externally developed early intervention. Operating on the assumption that all 11 of these reform schools might demonstrate similar achievement profiles, we included 3 typical comparison schools; these were schools with similar populations but with no history of either high achievement or reform activity. Two of the typical schools were in large urban districts and one was in a rural area. We wanted to include typical schools to provide a comparison base (both in terms of achievement and instructional practices) for the schools that had already undertaken and achieved some level of reform. All 3 of these schools were nominated by dis-

trict administrators as typical rather than above average for the district in terms of primary-grade students' reading achievement.

Table 1: Characteristics of Schools in the Study

School	Comm. Type	% Students on Subsidized Lunch	Grades	Type of Early Intervention	Type of School-wide Reform Program	Programs for English Language Learners	Classroom Teachers in Study	School Effectiveness Rating
1: Riverside	Rural	46	K-5	Nat'l grade 1: Individual	National		6	1
2: Stevenson	Rural	49	K-5	Local K-3: Groups and individual	Homegrown		8	3
3: Grant	Town	40	K-4	Reg. grades 1-2: Individual	Homegrown		6	2
4: Calvert	Town	66	K-5	None	National		8	2
5: Richland	Rural	28	K-5	National grade 1: Individual	None		8	1
6: Woodlawn	Suburb	38	K-5	Regular K-3: Group	Homegrown		7	3
7: Edison	City	56	K-5	None	None	ESL-Hmong	7	1
8: Lincoln	City	77	K-6	National grade 1: Individual	None	ESL-Hmong	8	2
9: Rosemont	City	85	K-3	National grade 1: Individual	Homegrown	Bilingual K-Spanish; ESL-Hmong, Spanish	8	2
10: Wheeler	City	92	K-2	Regular grades K-2: Group	Homegrown		4	3
11: Hilltop	Town	55	K-2	Regular grades K-2: Group	Homegrown	Bilingual ESL; K-2 Spanish	5	3
12: Colfax	Town	62	3-5	Regular grade 3: Group	Homegrown	Bilingual ESL; Grade 3 Spanish	2	2
13: Franklin	Rural	47	K-6	Regular grade K-3: Group	None	Bilingual and ESL; K-3 Spanish	8	1
14: Whitman	City	70	K-5	None	None	Bilingual and ESL; K-3 SPanish	6	2

Thus we began the study with 11 experimental schools and 3 control or comparison schools. However, as Stringfield et al. (1997) found in their work, not all schools believed to be exemplary in our study were, in fact, found to be so. Rather than rely on reputation, we decided to define school exemplarity empirically. We used a combination of gain scores from our own classroom reading measures plus scores on whatever achievement test the district normally used. Based upon this aggregate index, four schools in the present study were determined to be most effective and beating the odds. These schools were doing as well as or better than others in our sample in reading growth and/or doing better than average for their district, considering their poverty level. Six additional schools were determined to be moder-

ately effective (neither exceptionally high nor low on the two indices that made up our school effectiveness rating), and four schools were determined to be least effective—lower than other schools in our study on our composite index, but typical for their district in primary-grade reading achievement. While we did not begin this research with an eye toward comparisons across three levels of effectiveness among the schools we selected, it turned out that the natural variations within our sample permitted us to meet our original goal of examining systematic relations among those performance outcomes, program elements, and instructional practices.

Within each building in each of grades K–3, the principal was asked to identify two teachers whom he or she felt were good or excellent teachers and who would be willing to take part in the study. We did not include all teachers because we wanted to focus on exemplary practice. Similar to what we discovered for schools, however, not all teachers were found to be exemplary, at least according to the judgments of our expert ratings of teacher accomplishment. They varied widely along the scale of accomplishment that we used to characterize their practices. The principals contacted the teachers they had nominated to request their participation. In three schools, principals felt there was only 1 kindergarten teacher who met the criteria for inclusion in the study. In another school, only 1 teacher in kindergarten and 1 in grade 2 were identified by the principal for inclusion in the study. In one school, first grade did not participate because it was involved in another research study. In those schools with an early intervention or other form of supplemental instruction, one resource teacher was selected by the principal. We started in the fall with 107 teachers; a total of 3 teachers from two different schools withdrew from participation shortly after the study began. Thus, a total of 22 kindergarten, 23 first-grade, 25 second-grade, 22 third-grade, and 12 resource teachers participated in the study. All teachers were female except for 2 male second-grade and 2 male third-grade teachers. The number of teachers in each school at each grade level is included in Table 1.

Each principal was also asked to participate in the study. The principal recruited the teachers, responded to a survey, completed an interview, and provided demographic information about the school, including the number of students on free and reduced lunch, the school's overall performance on district tests, and grade 3 standardized test achievement (expressed as a percentile).

Our goal was to collect pre- and posttest data for four children per classroom—two average performers and two low performers. Mindful of problems with permission, attrition, and absences, we asked each teacher, in the fall, to select four average-achieving and four low-achieving children, based on teachers' perceptions of reading performance (or emergent literacy performance), to complete pretests. If more than four children per classroom remained as we prepared to collect posttest data, we used achievement level and gender balance to reduce the classroom pool to four students.

Basic Data Collection Tools

Fall and spring outcome measures

Children were pretested in November and again in May. With the exception of the writing and spelling assessments, which were often administered by the classroom teachers in small-group settings, all tests were administered by members of our research team who had been specially trained for this project and the administration of these tests.

Kindergarten. In the fall, children were individually tested on upper and lower case letter name identification (r = .96, alpha). In spring, the children were individually tested on upper and lower case letter name identification, rhyme (r = .88 alpha), phonemic blending and segmentation (r = .87, KR20, based on an earlier version of this test), and writing words. All kindergarten assessments came from the Emergent Literacy Survey (Pikulski, 1996), and none of the assessments was timed. For letter naming, the children were asked to name the upper case letters, presented in a scrambled order, and then name the lower case letters. For the phonemic segmentation and blending test, children were given an example of blending and an example of segmenting and then worked through six practice items. They were given the sounds in a word and asked to blend the sounds together into a word, or they were given a word and asked to give the sounds that were in the word. For the writing words test, they were asked to write as many words as they could. If needed, they were given the prompt, "Do you know your name, names of people in your family, names of animals, colors, any other words?" (Pikulski, 1996, p. 14).

Grade 1. In the fall, children were individually tested on upper and lower case letter name identification, phonemic blending, phonemic segmentation (Pikulski, 1996), and a list of preprimer words (described below). Children were asked to try to read a preprimer passage from the Qualitative Reading Inventory II (QRI-II) (Leslie & Caldwell, 1995). Because 83% could not yet read the preprimer passage with at least 90% accuracy, it was dropped from our analysis.

In the spring, children were individually assessed on a specially constructed word reading test and reading passages from the QRI-II. However, the normal QRI procedure was modified to accommodate our interests. Starting with the primer passage, students' word recognition accuracy score was recorded. If they could read the primer passage with at least 90% accuracy, they went on to the grade 1 passage and continued moving upward until instructional level (90% accuracy or better) was found. If they could not read the primer passage they were asked to read the preprimer passage. Once an instructional level was determined, each student, irrespective of decoding ability, was asked to read a grade 1 passage so we could obtain a common measure for all students—the number of words a child could read correctly in one minute (wcpm). This became our common metric of fluency (Deno, 1985). The QRI was administered by our data collectors who received training and had written directions to follow in the administration of the QRI-II. Two members of our team scored one third of the children's fall QRI passage from which the wcpm measure was taken (interrater reliability = 97% agreement).

Also, children were asked to retell each passage they read. A 4-point holistic scoring rubric, developed by the St. Vrain Valley School District in Longmont, CO (Colt, 1997), was used to score the retellings on the passage that proved to be at their instructional level, 90% accuracy (the retelling rubric appears in Appendix A). All of the retellings were scored by a single member of our research team. A second member read and scored 15% of the retellings to establish interrater reliability (91% agreement).

The reading words test was developed by the research team for the project in order to ensure that our word test included an appropriate mix of decodable words. Half of the words at each grade level were high frequency words from the QRI-II and half were decodable words garnered from an extensive review of the decodable patterns introduced in the four most popular basal series. The decodable words were controlled to match the QRI-II words in terms of frequency (Carroll, Davies, & Richman, 1971). There were 20 words at each grade level, preprimer through grade 3, for a total of 100 words. A child started with the preprimer list and continued until she made seven consecutive errors. The reading words test is included in Appendix B. The total reading words score exhibited a fairly high correlation with the words correct per minute measure (r = .82 in grade 1, .71 in grade 2, and .69 in grade 3); its split half (odd-even) reliability was .97.

Grade 2. In the fall, children were individually tested by a member of our research team on the reading words test for grade 1 and a grade 1 passage from the QRI-II. The number of words the child read correctly in the first minute was recorded. The child's word recognition accuracy on the passage was also recorded. Each child was asked to retell the passage, and the 4-point scoring rubric was used to score the retellings.

In the spring, children were individually tested by a member of the research team, starting with the reading words test. On the QRI-II, a procedure similar to the one used for grade 1 students was followed; in this case, each child began with a grade 1 passage and continued until an instructional level was found, after which each child read the grade 2 passage (to obtain the fluency measure) if it had not been read as a part of the procedure seeking to establish instructional level. For each passage, we asked the child for a retelling, and the retelling for the instructional level passage was scored using the St. Vrain 4-point rubric described earlier.

Grade 3. In the fall, children were individually tested on reading words (grade 2 list) and on a grade 2 passage from the QRI-II. They were asked to read and retell the passage; from these data, three scores were computed: word recognition accuracy, fluency (wcpm), and retelling.

In the spring, third graders followed the same general procedure as second graders. They were individually tested on the reading words test; beginning with the preprimer list, they read successive lists until they missed seven consecutive words. They read passages from the QRI-II, starting with the grade 2 passage and continuing, either up or down, until an instructional level was found, at least through the grade 4 passage. For each passage read, a child's word recognition score and passage-retelling score was recorded. If it had not been encountered in the procedure to establish instructional level, every third grader was asked to read from the grade 3 passage so that a

fluency measure (wcpm) could be obtained for every child on a grade-level passage.

Other measures

A writing sample based upon a common prompt and a graded spelling test, adapted from the work of Viisc (1994), were also administered in the fall for students in grades 2-3 and in the spring for all students in grades 1-3. Typically these were administered by the classroom teacher or another staff member at the school. An examination of these samples revealed such inconsistency in administration and such low levels of performance that they were dropped from any further analysis.

Observations

Members of our research team at each site were trained to conduct classroom observations. Observations occurred during an hour in which the teacher provided reading instruction. The observer recorded what the teacher was saying and doing as well as what the children were saying and doing during the lesson. Every five minutes the observer recorded any of the following teacher behaviors which were observed in the previous five-minute segment: coaching/scaffolding, modeling, engaging the children in recitation, explaining how to do something, telling, or engaging the children in a discussion. A description of each of these behaviors is provided in Table 2. Observers practiced coding video segments of instruction until they had at least 80% agreement with the principal investigators on the coding of teacher behaviors.

Table 2: Teachers' Interaction Styles

Style	Description
Coaching/scaffolding	Teacher support/prompts/coaches child as he or she tries to perform a task
Modeling/demonstrating	Teacher shows how to do something by actually performing a task or thinking aloud
Engaging students in recitation	Teacher engages children in a questioning session in which he or she is looking for short, specific answers
Telling students information	Teacher gives facts in the abstract without directions on how to use the information or what to do with it
Explaining how to do something	Teacher provides direct explanation of processes involved in a task
Engaging students in discussion	Teacher leads students in a formal discussion in which the social conventions of a discussion apply

During the observations, the observers primarily focused on the teacher and on one low and one average target child (each of whom had been involved in the testing). However, comments on other children were included as well, and observers also provided comments about general classroom activity, involvement in the lesson, and other activities that seemed noteworthy. After the observation, the observer summarized the lesson by completing a summary form that required a statement about each of these characteristics: overall impression, teacher instruction and teacher-student interaction, activities and materials, student engagement, classroom management, and classroom environment.

Beginning in December, a member of our team conducted a one-hour observation of instruction during the basic reading program in every classroom once a month for five months. Observations were scheduled to accommodate each teacher's schedule. Complete observational data were obtained for 92 classroom teachers.

Logs

We asked every classroom teacher to keep a log of daily instructional activities in the classroom for one week in February and one week in April. We asked them to indicate how long they spent on various activities, including teacher directed reading of narrative and expository text; student independent reading; instruction in phonics, vocabulary, and comprehension; literature circles; writing in response to reading; other written composition; spelling; reading aloud to students; and other academic activities. Teachers recorded activities in 15-minute intervals and could include more than one activity during a time period. We divided the number of minutes for an interval by the number of activities coded to get number of minutes spent on an activity during that interval. For example, if a 15-minute interval was coded as whole-group phonics instruction, independent reading, and writing in response to reading, we coded each activity as occurring for a child for 5 minutes. Teachers also indicated the group setting in which each activity occurred: students working as a whole group, working in a small group, or working independently.

Table 3: Indicators of Instructional Emphasis

Activity Grouping	1: Reading Instruction	2: Independent Reading	3: Writing in Response to Reading	4: Total Reading	5: Reading Aloud	6: Composition	7: Spelling	8: Total
Large	X							
Small	X							
Independent	X							
Total	X	X	X	X	X	X	X	X

Using the raw data from the logs, we created three setting codes and six basic activity codes. The setting codes (e.g., large group, small group, and one-on-one) were fairly transparent and easy to translate into categories for further analysis. The original activity codes (e.g., phonics, comprehension, vocabulary, directed reading of narrative text, directed reading of informational text, literature circles) proved somewhat more challenging because of the ways in which teachers interpreted the directions. Thus we decided to create one large category labeled reading instruction (which included teacher-directed reading of narrative and expository text; instruction in phonics, vocabulary, and comprehension; and literature circles); the other original categories—independent reading, writing in response to reading, other written composition, spelling, reading aloud to students—remained intact. This data set allowed us to create a large number of different indicators of instructional emphasis. From those possibilities, presented in chart form in Table 3, we chose those variables in which the cells are marked with an X. In other words, we selected each of the activities summed over settings, each of the settings summed over activities, and a composite variable for total reading emphasis, which is the sum of the totals of columns 1-3

(reading instruction, independent reading, and writing in response to reading).

Questionnaires

In April or May, the principal and teachers from each school completed surveys which had been developed by a team of CIERA researchers for a broader national survey of beat the odds schools. The principal survey dealt with the following topics:

- Use of goals and standards;
- factors in improving reading/language arts instruction;
- approaches to professional development;
- planning reading/language arts instruction;
- attributes of effective schools;
- attributes of effective instruction;
- supplemental programs;
- role of special education teachers in reading/language arts instruction;
- perceptions of programs for struggling readers;
- increasing achievement for struggling readers;
- use of computers in reading/language arts;
- showcases for student work;
- consequences of assessment;
- assessment training; and
- reasons the school in question was successful in teaching students to read.

The teacher survey embraced most of the same topics and some additional ones:

- use of curriculum materials;
- aspirations for student reading levels tied to familiar trade books;
- time spent on reading/language arts activities;
- types, frequencies, and purposes of assessments;
- types of reading/language arts helpers and activities they engaged in;
- home-school communication and activities; and
- community activities.

Although the surveys were long (16 pages), the return rate was high. Across the 14 schools, 93 teacher surveys and 11 principal surveys were returned

from the 104 teachers and 14 principals who participated in the study, for a return rate of 88%.

Interviews

We interviewed all principals and at least three teachers from each school. Principal questions focused on the community and links to parents, the principal's view of his or her leadership role, factors contributing to the school's success, challenges as well as things on which the school was still working, and advice to schools that wanted to significantly improve their reading achievement. Teacher questions were similar but also included questions about a teacher's general approach to teaching reading, behavior management systems, and her or his expectations for students. We transcribed the interviews and used them as a source of information for writing case studies for each school and for generating several school variables.

Table 4a: Descriptions of School Factors

Definition	Description	Sources of Data
1. School effectiveness rating: Based on three project reading measures and mean grade 3 percentile on district standardized test	Composite *z*-score from: 1. Composite *z*-score on three project reading measures: spring reading words residual, spring retelling at reading level residual, and spring words-correct-per-minute residual 2. *Z*-score from residual on third grade standardized reading text percentile (poverty used as a covariate)	One half of rating from three project reading measures and one half from district standardized reading test
2. Parent links: Based on teachers' mean level of home communication and on efforts of school to reach out to parents	Rating with three levels: 3-High-high on parent communication and high on involving parents 2-Average-One or the other of the above 1-Low-None of the above or low on parent communication (see teacher factor)	Teacher survey, case studies, and interviews
3. Systematic evaluation of pupil progress: Regular monitoring of reading progress, used by all teachers, results shared, data used to make instructional changes	Rating with two levels: 2-Present 1-Absent	Teacher surveys, case studies, and interviews
4. Building communication: Based on teachers' rating of "communication of ideas across teachers" (from survey) as indicative (5) down to absent (1) in school; also based on positive or negative comments in interviews, case study about communication and collaboration across teachers	Rating with three levels: 3-High-High on teacher's perceptions of building communication and positive comments in interviews/case study 2-Average-Average teacher perceptions 1-Low-Low teacher perceptions and negative comments in interviews/case study	Teacher surveys, case studies, and interviews

Case studies

For every school, a common outline was followed to guide our research team in crafting a case study for each school. A model case study was also provided to create some common expectations for content, format, and depth. Major topics within each case study included:

1. School demographics;
2. history of the school;
3. primary-grade reading performance data;
4. community/home/school relationships;

5. school factors (leadership, effective schools/school change factors, professional development, format of school reading program);
6. grade-level classroom practices (instruction, curriculum, student engagement, classroom management, classroom environment);
7. other factors believed to be enhancing beginning reading achievement;
8. challenges; and
9. advice to other schools.

Table 4b: Descriptions of Teacher Factors

Definition	Description	Sources of Data
1. Home communication: Based on teachers' report on survey	Rating with five levels: 5-Very high-Exceeded rating of 4 4-High-Called home at least once a month and sent letter, newsletter, or traveling folder (i.e., artifacts) home 2+ times per week 3-Average-Called home once a month and sent 2+ artifacts home per month *or* called home 2-3 times per year and sent home 1 artifact per week 2-Lower than average-Called home 2-3 times a year and sent home 1 artifact per month *or* called home once a year and sent home 1 artifact per week 1-Low-Less than required for rating of 2	Teacher surveys
2. Student engagement: Based on comments in observations about student engagement	Rating with three levels: 3-High-high in maintaining engagement; most students engaged (in most comments in observations) 2-Average-Average in maintaining engagement; some comments indicated high engagement and others indicated a number of students off-task 1-Low-Low in maintaining engagement; many comments indicated many students often off-task	Observation notes and observation summaries
3. Time spent in reading instruction and activities	Activities included: whole-group, small-group, or independent grouping pattern on the following activities: reading instruction (including teacher directed reading, word recognition, comprehension, vocabulary instruction, or literature circle); independent reading; writing in response to reading; reading aloud to students; written composition; spelling; and other academic activities	Teacher logs
4. Preferred interaction style: Based on most frequent behavior coded in observations	Six behaviors coded: 1-coaching/scaffolding; 2-modeling; 3-engaging students in recitation; 4-explaining how to do something; 5-telling students information; 6-engaging student in discussion	Observation notes
5. Approaches to word recognition	1. Coaching on use of word recognition strategies during reading of stories 2. Providing explicit phonics instruction 3. Practicing sight words	Observation notes
6. Approaches to comprehension instruction	1. Asking text-based questions to be answered orally after reading 2. Asking higher level questions to be answered orally after reading 3. Writing in response to reading	Observation notes
7. Teacher accomplishment rating: Based on two experts reading observations	Rating with three levels: 3-Demonstrating many of the elements of culturally responsive teaching and elements of effective instruction 2-Demonstrating some of the elements of culturally responsive teaching and elements of effective instruction 1-Demonstrating few of the elements of culturally responsive teaching and elements of effective instruction	Observation notes and summaries

Case studies were written by the principal investigator or by the research team member who had spent the most time in a particular school. They ranged from 15–30 pages each.

Creating Variables for Data Analyses

Overall strategy

Quantitative and descriptive analyses were conducted using multiple sources of information used for both types of analysis. Analyses were conducted at the school and at the classroom level. As indicated, we made a decision fairly early on that we stood to learn the most from this initiative by examining the natural variability within our sample. As a general strategy, then, we developed empirically driven indices of effectiveness for both schools and teachers, grouped the schools or teachers into categories of effectiveness, and then examined the systematic differences among schools or teachers on other sets of variables. The procedures and criteria used to determine these levels will be described later. In addition, we developed categories and rating systems, built from data emerging from the observations, surveys, and interviews, that enabled us to operationalize both school-level and classroom-level variables. Tables 4a and 4b provide a summary of the school and teacher variables, including the procedures used to construct them.

Student outcomes

In order to examine the effects of the organizational and instructional factors that we observed in these schools, it was necessary to create outcome measures that reflected the effects that might conceivably have occurred during our year of observation. Therefore we decided to create residual scores for all the relevant spring reading measures, using appropriate fall scores as covariates. In grade 1, spring residual scores for reading fluency (wcpm) and retelling were created by using fall phonemic segmentation and blending scores as covariates. In grades 1–3, spring residual scores for reading words were created by using fall reading words scores as a covariate. In grades 2–3, spring residual scores for reading fluency and retellings were created by using fall reading fluency scores as a covariate. Within each grade level, these residual scores were converted to *z*-scores, which were calculated from the mean and standard deviation for the entire grade level sample, so that the data could be aggregated and/or compared across grade levels.

Constructing school variables

School effectiveness rating. To be able to categorize schools as most effective, moderately effective, or least effective, we used two school-based measures: (a) students' growth on project measures of reading, and (b) students' performance on district measures of reading. These two scores were combined to create a general rating of effectiveness. First, the residual *z*-scores for retelling (at instructional level), fluency (wcpm on a grade level passage), and reading words were aggregated and standardized to create a composite index of reading growth. Second, we calculated what might be called a primary (as in primary grades) outcome index, using the end-of-grade-3 scores on the district-mandated test. As it turned out, these were, in each case, standardized achievement tests (six schools used the Stanford Achievement Test 9; two used the Metropolitan Achievement Test 7; two used the California Achievement Test; two used the Northwest Evaluation Association Levels Test; and two used a district-normed test.) A residual

mean percentile score for each school was calculated by controlling for the school's poverty level (as indexed by the percentage of students receiving free or reduced lunch). This was done because students' achievement scores are depressed in schools with 50–75% of students living in poverty and seriously depressed in schools with 75–100% of students living in poverty (Puma et al., 1997). The residual scores were then converted to *z*-scores.

Third, the *z*-scores on the project measures and primary grades outcome measure were summed and standardized. When we examined these scores, we looked for natural breaks in the distribution that would divide the schools into three groups of approximately the same size. These breaks occurred at .5 standard deviations above and below the mean. Breaking at those two points yielded four most effective, six moderately effective, and four least effective schools (see Table 5). As the data in Table 5 suggest, the most effective schools were not necessarily populated by the most economically advantaged students.

Table 5: *Z*-scores for School Effectiveness Rating

School	Standard Composite *Z*-score From *Z*-scores (by Grade) for Residual WCPM*, Residual Retelling, Residual Reading Words	*Z*-score for Residual Mean Percentile on District Standardized Reading Test (Controlling for Poverty)	Sum of *Z*-scores (Columns 2 & 3)	School Effectiveness Score—Standardized Score (From Column 4)	School Effectiveness Rating 3-level†
1	-.74	-.22	-.96	-.58	1
2	.74	.53	1.27	.77	3
3	-.22	-.36	-.58	-.35	2
4	-.01	-.52	-.53	-.32	2
5	.03	-.91	-.88	-.53	1
6	1.33	2.04	3.37	2.04	3
7	-1.39	-1.28	-2.67	-1.62	1
8	-.41	1.17	.76	.46	2
9	1.16	-.56	.60	.36	2
10	1.74	-.39	1.38	.82	3
11	.28	1.45	1.73	1.05	3
12	-.39	-.36	-.75	-.45	2
13	-1.91	-.75	-2.56	-1.56	1
14	-.32	.15	-.17	-.10	2

* Words correct per minute

† 3 (Most effective if > .50 *SD* above mean of 0)
2 (Moderately effective if < .50 *SD* above mean but > -.50 *SD* below mean of 0)
1 (Least effective if < -.50 *SD* below mean of 0).

School efforts to link to parents. An initial reading of the case studies indicated that schools varied in the extent to which they reached out to parents. Schools were judged to be high on the linking to parents factor if they had (a) a high mean score on teacher's level of home communication

(described below under teacher factors), and (b) one or more of the following in place:

1. An active site council in which parents helped to make school decisions;
2. focus groups, phone surveys, or written surveys to find out parents' needs and concerns;
3. family education programs on site in which meals and day care were provided; or
4. a principal who called parents regularly to deliver good news about their children.

Schools were judged to be average on the linking to parents measure if they had one of the above factors (a or b) present in their school. Schools were judged to be low on the linking to parents measure if they demonstrated neither of the above factors or if they were low on the home communication teacher factor. Six schools were determined to be high on the linking to parents school factor, five were determined to be average on this factor, and three schools were determined to be low. The two raters evaluating schools on this factor achieved 93% agreement in their categorizations.

Systematic, internal assessment of pupil progress. A school was coded as systematically assessing pupil progress if at least two thirds of teachers on the survey perceived this to be an attribute of their school, and if comments in the case studies and/or interviews supported this perception. Across schools, all but four were coded as systematically assessing pupil progress. These were not externally imposed standardized testing systems; they were internally developed systems for monitoring individual student progress within a schoolwide curriculum.

Building communication. In a number of the interviews and surveys, staff commented about either good or poor collaboration and/or communication among teachers. Schools were judged to be high on the building communication scale if they had the following present: a high mean score on teachers' perceived building communication rating (from the teacher survey) and positive comments in the case study or in interviews about good collaboration among teachers within and across grades. Schools were judged to be average on this building communication factor if they had (a) an average mean score on the teachers' perceived building communication rating and no negative comments in the case study or interviews about building level communication or (b) a high mean on the teacher rating but negative comments about building level communication in the case study or interviews. A school was judged to be low on the building level communication factor if the teachers' mean building communication rating was low and negative comments about building level communication appeared in the case study or interviews. Two raters achieved 86% agreement on this judgment. Across the 14 schools in the project, 6 schools were judged to be high on the building communication factor, 5 were judged to be average, and 3 were judged to be low.

Use of an externally developed early reading intervention. Within each grade level, a school was coded as either having an externally devel-

oped early intervention or no externally developed intervention in place. Out of the 14 schools, 10 had an externally developed intervention, 5 across two or more grades, 4 in grade 1 only, and 1 in grade 3 (in a grade 3–6 building.) Four schools had no externally-developed intervention in place.

Constructing classroom variables

Home communication. On the survey, teachers indicated how often they communicated with parents and in what ways. The areas in which there appeared to be the most variability were the frequency with which teachers reported (a) calling home, (b) sending a letter or newsletter home, or (c) sending a traveling folder home. A 5-point scale was used to rate teachers on the extent to which they communicated with parents. (In the list below, letters, newsletters, and folders are all grouped together as artifacts.)

5. (very high) Exceeds standards for level 4

4. (high) Calls home at least once per month and sends home 2+ artifacts per week

3. (moderate) Calls home once per month and sends home 2 artifacts per month or calls home 2–3 times per year and sends home 1 artifact per week

2. (low) Calls home 2–3 times per year and sends home 1 artifact per month or calls home 1 time per year and sends home 1 artifact per week

1. (very low) Does not meet standards for level 2

Across the sample of teachers in this project, 19% received a rating of 5, 19% a rating of 4, 40% a rating of 3, 14% a rating of 2, and 8% a rating of 1. The mean rating was 3.28 (SD = 1.16). One research team member rated all teachers, and a second rated a 25% sample; interrater agreement was 95%.

Student engagement. From the observations, we searched for comments about children's engagement, which for us embraced both compliance (on-task behavior, in which children are productively engaged in their assigned activity) and involvement (evidence of genuine student enthusiasm for the activities). Most helpful in this regard were the summaries that each observer completed at the end of each observation session; one of their tasks was to summarize and point to evidence to support any conclusions about engagement. A teacher received a high rating (3) in maintaining student engagement if most of the comments on either of these topics (compliance or involvement) indicated that most students were engaged. A teacher was coded as average (2) if, in looking across comments about engagement, the comments were mixed—some indicating high student engagement and others indicating low engagement and/or off-task behavior. A teacher received a low (1) rating if numerous comments indicated that many students were off task. One member of our research team rated all teachers and another team member rated 25% of the teachers on this dimension. Across the pairs of ratings for each teacher, there was 100% agreement. Across teachers, the mean rating was 2.30 (SD = .76), with 48% of teachers receiving a rating of 3 (high), 34% a rating of 2 (average), and 18% a rating of 1 (low) in maintaining student engagement.

Preferred interaction style. At the end of each five-minute segment during the classroom observations, observers coded instances of interactions

observed during that segment, using these categories: coaching/scaffolding, modeling, engaging students in recitation, engaging students in discussion, explaining how to do something, or telling students information. (These activities are described in more detail in Table 3.) The total number of times a teacher was coded as engaging in each of these behaviors (more than one was possible within each segment) was calculated, and the behavior coded most frequently across all five observations was determined to be the teacher's preferred interaction style. Across teachers in grades 1, 2, and 3, 24% had a preferred interaction style of coaching, 31% had a preferred interaction style of engaging students in recitation, 39% had a preferred interaction style of telling students information, and 6% had a preferred style of modeling. To evaluate the trustworthiness of these ratings, three members of the research team rated a total of 25% of the teachers in terms of preferred interaction style based on a reading of the classroom observations and observation summaries for these teachers. There was 82% interrater agreement with the preferred interaction style as determined from the observation codes of interaction styles at five-minute intervals. If and when mismatches occurred, we reverted to code most frequently marked by the classroom observer.

Approaches to word recognition and comprehension instruction. The data from the observations were analyzed to determine how teachers provided word recognition and comprehension instruction. A number of approaches were coded on a frequency scale, where each approach was determined to occur frequently, occasionally, or never during our observations for a particular teacher. The word recognition approaches included (a) coaching children in the use of strategies to figure out unknown words as they were reading text, (b) focusing on words in stories to review phonic elements, (c) providing explicit phonics instruction, and (d) practicing sight words.

Coaching involved prompting children to use a variety of strategies as they were engaged in reading during small-group instruction or one-on-one reading time. Typical prompts included phrases such as:

- What do you do when you come to a word you don't know?
- Does that make sense? (e.g., *tease* for *teach*)
- What do the letters *ea* say?
- How do you know it can't be *dripped*? (for *dressed*)
- Do you see a chunk you recognize? (in *nothing*)
- Let's sound it out and think of what would make sense
- Let's reread like good readers do.

Focusing on words in stories was realized most often as asking children to frame a word as they were reading in order to highlight a particular phonic element. For example, one teacher told the children to point to *feet* in the title of a new story and told the group that the ee said /e/ (long *e* sound). Another approach encouraged children to search for a particular element in their stories after reading. One teacher who was talking about *-ed* endings asked children to look for words in their story with *-ed* endings. Only four

teachers in the entire sample were frequently observed using this approach, so it was eliminated from further analysis.

Explicit phonics instruction included work on a chart, whiteboard, worksheet, or word cards dealing with word study; word families; introducing or comparing phonic elements (i.e., *er, ir,* and *ur* all have the same sound); making words (Cunningham & Cunningham, 1992); writing words; and reading words with a particular phonic element in isolation. Practice on sight words involved teachers using flash cards, a pocket chart, or a word wall to review words the students were expected to recognize instantly as sight words.

For comprehension instruction, eight different instructional practices were observed and coded: doing a picture walk; asking for a prediction; asking a text-based question; asking a higher level, aesthetic response question; asking children to write in response to reading (including writing answers to questions about what they had read); doing a story map; asking children to retell a story; and working on a comprehension skill or strategy. Each grade 1–3 teacher was coded as frequently, occasionally, or never observed engaging in each of the above practices. For further analysis, we focused on those categories for which 10 or more teachers were frequently observed to have used the strategy: asking text-based questions, asking higher level questions, and asking children to write in response to what they had read.

One member of the research team coded all teachers and a second member coded 25% of the teachers. There was 100% agreement on coding of coaching for word recognition during reading, 94% agreement on providing explicit phonics instruction, 94% agreement on practicing sight words, 96% agreement on asking text-based questions, 96% agreement on asking higher level questions, and 100% agreement on writing in response to reading.

Teacher accomplishment rating. Two experts in teacher supervision at the elementary school level, both members of our research team, read all teacher observations. One expert was a whole language advocate, and the other described herself as a reading skills advocate. They used a checklist of elements of effective instruction from the principal survey (based on Anderson et al., 1979; Barr & Dreeban, 1991; Hoffman, 1991; Pressley, Rankin, & Yokoi, 1996; Roehler & Duffy, 1991; Wharton-MacDonald et al., 1998) and a checklist of elements of culturally responsive teaching from the work of Ladson-Billings (1994; see Table 6). Although the two lists have a fair amount of overlap, they also tap different aspects of skillful teaching.

For each item in each list the raters estimated, based on a thorough reading of the observations and summaries, the quantity of indicators of demonstrated accomplishment by a teacher: many, some, few, or could not determine. Each rater rated each teacher separately on the two scales. For each teacher, the raters examined each item on each scale separately, providing a judgment about whether there were many, some, or few instances of that attribute in the data set (or if there were data, they coded the item as *could not determine*. In 98% of the cases, a teacher received the same score on both scales. The two raters then arrived at a composite rating (on a 3-point scale) based upon a mean of the two scores. For the overall rating, the two raters demonstrated 80% agreement on their overall ratings. The two raters had used a system of pluses and minuses along with their numerical ratings. When ratings less than a point apart (e.g., 2-, 1+; 3, 2+) were considered, the

Table 6: Two Sets of Criteria Used to Rate Teacher Accomplishment

Elements of Culturally Responsive Teaching	Elements of Effective Instruction
1. Seeing teaching as art	1. Awareness of purpose
2. Seeing oneself as part of the community	2. Enthusiasm
3. Believing all students can succeed	3. Task orientation
4. Helping students make connections between community, nation, world, and self	4. High pupil engagement
5. Teaching from the perspective of "pulling knowledge out" instead of "putting knowledge in"	5. Short transitions
6. Viewing teacher-student interactions as fluid, humanely equitable, extending to interactions beyond the classroom	6. Much time spent in reading/ language arts activities
7. Demonstrating a connectedness with all students	7. Frequent instruction in skills and strategies
8. Encouraging a community of learners	8. High success rate
9. Encouraging students to learn collaboratively	9. Masterful classroom management
10. Viewing knowledge as being recreated and shared by teachers and students	10. Positive classroom climate
11. Viewing knowledge critically	11. High pupil expectations
12. Teaching content with passion	12. Redoubling of teaching efforts when students have difficulty
13. Helping students develop necessary skills	13. Effective use of praise
14. Seeing excellence as complex but taking account of student diversity and individual differences	14. Extensive content converage, instructional density
15. Treating student as competent	15. Explicit modeling and scaffolding
16. Providing instructional scaffolding	16. Teaching skills in context
17. Making instruction the focus of the classroom	17. Extra instruction for low readers
18. Extending students' thinkng	18. Encouragement of self-regulation
19. Possessing in-depth knowledge about the students and the subject matter	19. Instructional balance
	20. Much reading of connected text
	21. Much writing of connected text
	22. Activities appropriate, meaningful, challenging

two raters had 94% agreement. There were only six instances in which the raters disagreed by a point, and these disagreements were moderated by consultation with the senior author of this paper. Across teachers in our sample, 41% were identified as demonstrating many of these elements (most accomplished), 32% demonstrated some of these elements (moderately accomplished), and 27% demonstrated few of the elements of culturally responsive teaching and effective instruction (least accomplished).

Results and Discussion

The results are organized by level of analysis. First we report and discuss results, largely descriptive, at the school level. Second, we examine the variations among instructional practices of teachers within the levels of school effectiveness that emerge from the school level analysis. Third, we report and discuss the variations we found in instructional practice as a function of levels of teacher accomplishment (how they were rated on the two-pronged scale—effective teaching and culturally responsive pedagogy). Finally, we

report on a regression analysis combining school and teacher variables. Throughout the results we refer to the most effective schools by name: Hilltop, Stevenson, Wheeler, and Woodlawn (all pseudonyms).

Due to the complexity of the study, the fact that many of the classroom variables focus on grades 1-3 (e.g., student level of engagement, time spent in small- or whole-group instruction, preferred interaction style), and the use of different outcome measures, the kindergarten classrooms were dropped from the analysis. A brief summary of findings are in Appendix C.

Table 7: Grade 1 Means (and Standard Deviations) for Pre- and Posttest Scores by School Effectiveness Rating and Teacher Accomplishment Rating

	School	N*	Fall			Spring		
			Letter ID	Phonemic Awareness	Preprimer Words	Reading Words	WCPM† on Grade 1 Passage	Retelling at Reading Level
School rating‡	1	8	47.95 (3.51)	9.38 (1.71)	9.90 (4.58)	56.28 (14.40)	34.68 (12.09)	2.52 (.64)
	2	8	44.75 (5.82)	6.56 (3.85)	7.38 (4.44)	55.87 (16.89)	41.19 (14.99)	2.34 (.30)
	3	7	45.80 (3.97)	7.48 (2.51)	9.94 (3.87)	68.29 (9.01)	50.75 (17.03)	2.51 (.35)
Teacher rating**	1	3	43.00 (1.64)	4.52 (4.68)	4.58 (2.18)	41.58 (14.62)	25.58 (11.22)	2.17 (.14)
	2	7	45.83 (4.42)	7.11 (1.96)	7.67 (3.80)	66.18 (17.70)	44.40 (15.96)	2.38 (.37)
	3	12	46.74 (5.13)	8.57 (2.87	10.67 (4.23)	60.48 (10.32)	44.44 (15.39)	2.59 (.52)

* Number of teachers.
† Words correct per minute.
‡ 1 = Least effective; 2 = moderately effective; 3 = most effective.
** 1 = Least accomplished; 2 = moderately accomplished; 3 = most accomplished.

The unit of analysis depended on the level. Some analyses involving school factors were based on the school-level scores (e.g., school-level ratings, composite school effectiveness scores that had been summed across grade level) for each of the 14 schools. Most analyses at the school and all at the teacher level were based on classroom mean scores. Mean classroom performance measures by school effectiveness and teacher accomplishment levels are reported, for archival purposes, in Tables 7-9.

School Level Analyses

Linking to parents

Parent links were positively and statistically significantly related (r = .73) to the school effectiveness rating and to all measures of student growth, fluency (r = .60), retelling (r = .37), and reading words (r = .41; see Table 10).

Table 8: Grade 2 Means (and Standard Deviations) for Pre- and Posttest Scores by School Effectiveness Rating and Teacher Accomplishment Rating

	School	N*	Fall		Spring		
			Grade 1 Words	WCPM† on Grade 1 Text	Reading Words	WCPM† on Grade 2 Text	Retelling at Reading Level
School rating‡	1	8	13.88 (3.59)	50.46 (18.92)	79.59 (5.88)	66.30 (19.24)	2.56 (.69)
	2	10	15.88 (2.18)	59.77 (13.07)	86.98 (6.04)	73.35 (14.65)	2.99 (.36)
	3	6	14.42 (3.46)	47.98 (17.31)	84.58 (7.99)	70.17 (13.18)	3.06 (.40)
Teacher rating**	1	8	13.91 (2.50)	56.83 (13.38)	80.22 (8.09)	64.47 (17.37)	2.66 (.78)
	2	8	15.20 (3.58)	55.63 (14.38)	84.82 (7.00)	71.40 (17.20)	2.99 (.39)
	3	8	15.44 (3.12)	48.68 (21.31)	86.72 (4.73)	74.75 (11.80)	2.95 (.30)

* Number of teachers.

† Words correct per minute.

‡ 1 = Least effective; 2 = moderately effective; 3 = most effective.

** 1 = Least accomplished; 2 = moderately accomplished; 3 = most accomplished.

The most effective schools reported more links with parents than the moderately effective and least effective schools. Three of the four most effective schools reported having an active site council in which parents served on the committee with teachers and other school staff and helped to make

Table 9: Grade 3 Means (and Standard Deviations) for Pre- and Posttest Scores by School Effectiveness Rating and Teacher Accomplishment Rating

	School	N*	Fall		Spring		
			Grade 2 Words	WCPM† on Grade 2 Passage	Reading Words	WCPM† on Grade 3 Passage	Retelling at Reading Level
School rating‡	1	7	16.95 (2.15)	73.19 (14.48)	90.99 (3.76)	85.70 (18.98)	2.80 (.29)
	2	2	15.27 (3.74)	75.58 (22.79)	89.89 (8.14)	91.13 (29.13)	3.06 (.43)
	3	3	16.81 (2.03)	72.58 (8.82)	96.25 (1.17)	105.38 (16.92)	3.31 (.47)
Teacher rating**	1	5	15.95 (2.27)	71.70 (11.67)	90.60 (3.91)	88.49 (27.83)	2.96 (.38)
	2	9	16.28 (3.67)	75.88 (17.37)	91.44 (8.21)	92.08 (23.59)	3.14 (.50)
	3	7	16.20 (3.23)	77.11 (23.06)	92.33 (46.48)	94.58 (31.35)	2.92 (.35)

* Number of teachers.

† Words correct per minute.

‡ 1 = Least effective; 2 = moderately effective; 3 = most effective.

** 1 = Least accomplished; 2 = moderately accomplished; 3 = most accomplished.

decisions concerning school practices. Four of six moderately successful schools also reported having an active site council, but only one of the four least successful schools reported having such a body in place.

Additionally, the most effective schools reached out to parents in other ways. Wheeler scheduled focus groups to learn how to better meet parents' and students' needs. Woodlawn had regular focus groups and conducted phone surveys to determine parents' concerns and needs. Stevenson school officials sent a written survey home to parents, and the principal attempted to communicate with all parents regularly. None of the moderately successful or least successful schools reported engaging in any of these practices.

Hilltop, another of the most effective schools, had developed an at-home reading partnership in which books were sent home in English or Spanish so that parents could read to their children. Woodlawn and Stevenson also cited successful at-home reading partnerships, all with high parent participation, in which parents listened to their children read on a regular basis.

When asked about reasons for their success, Stevenson and Hilltop mentioned the importance of strong home-school connections, reaching out to parents and making them feel like welcome partners in the school. One teacher at Stevenson conveyed the common sentiment succinctly: "One factor responsible for our success is working closely with the home. I include parents in as many ways as possible—send a letter explaining a unit, send homework tips. I try to communicate to parents the power they have in influencing their students' growth."

Table 10: Correlations for School Effectiveness Rating and School Factors

	School Effectiveness Score	Z-score Residual Words (Class Mean)	Z-score Residual Fluency (Class Mean)	Z-score Residual Retelling (Class Mean)
Links to parents	.73 $p < .01$ $n = 14$	.41 $p < .001$ $n = 70$	.60 $p < .001$ $n = 70$	.37 $p < .01$ $n = 70$
Systematic assessment of pupil progress	.42 $p > .05$ $n = 14$	.21 $p > .05$ $n = 70$	.53 $p < .01$ $n = 70$	.37 $p < .01$ $n = 70$
Building communication	.37 $p > .05$ $n = 14$	.07 $p > .05$ $n = 70$	.43 $p < .001$ $n = 70$	.35 $p < .01$ $n = 70$
Research-based intervention in place	.00 *ns* $n = 14$	.14 *ns* $n = 70$	.02 *ns* $n = 70$	.05 *ns* $n = 70$

While realizing that correlation does not imply causation, we conclude that part of positive home-school relations, so often cited in the literature and so prevalent in our most effective schools, is making a concerted effort to reach out to parents. As an example of this concerted effort, the principal at Stevenson, although perhaps an exception in this respect, made time herself to call all homes just to compliment parents on the academic or social achievements of their children.

Systematic internal assessment

Another of the eight attributes of effective schools discussed by Hoffman (1991), systematic assessment of pupil progress, figured prominently in our findings. Systematic assessment was statistifically significantly correlated with students' growth in reading fluency (r = .53) and with their retelling performance (r =.37).All four of these schools used some form of systematic assessment of student progress. In all four cases, this meant that all the teachers in the school regularly (at least three times throughout the year) administered some sort of common classroom-based assessment tool to all students and shared the information about classroom-level performance with the principal and fellow teachers. Wheeler implemented four assessments across the school year, relying on a fluency measure (wcpm), sight words, and letter identification in grades 1 and 2, and letter identification and concepts about print in kindergarten. Woodlawn used a fluency measure as a curriculum-based indicator five times a year (three times in first grade). Hilltop used an informal reading inventory three times a year. Stevenson used an informal inventory (grade 1) or basal tests (grade 1-2) three times a year as well as a developmental spelling test three times a year in grades 1-3 and a words-in-isolation-and-in-context test in grade 1 every six weeks. Four of the six moderately effective schools also reported having systematic assessment of pupil progress in place, while two of the four least effective schools had an assessment system in place.

We would emphasize that these were curriculum-based, classroom assessments intended to provide information for monitoring individual student progress and to shape individual and classroom (and occasionally schoolwide) curricular and instructional decisions; they were not external, accountability-focused assessments. Instead of external accountability, these classroom-level data provided a form of internal accountability (to one's colleagues) while providing teachers with a useful benchmark on each student's progress. The public sharing of the data was important in all four schools. The principal at Hilltop related that the teachers had learned how to confront data, to keep data in front of them, to use data to identify specific strategies to help struggling readers, to provide support in the implementation of strategies, and to align major school events and celebrations around the meeting of schoolwide goals. In other words, the staff at this particular school had learned how to make performance data a useful ally rather than a cause of constant alarm or frustration.

Building communication

Building communication was positively related to the fluency measure (r = .43) and retelling (r = .35). Five of the ten moderately effective or least effective schools revealed concerns about communication across teachers and program articulation across grades. The interviews with staff in one moderately and one least effective school revealed several instances of negative communication and collaboration, including low morale among teachers due to the existence of different factions among the staff and perceived lack of cooperation among teachers. In one of the least effective schools, all of the teachers rated communication of ideas across teachers as moderate to low in their school. In another least effective school, working as a team across the entire school was highlighted as something the school had just begun to work on; seven of nine teachers in that school rated the presence of building communication as moderate to low in their school.

Teachers in all four of the most effective schools reported collaboration within and across grades as a reason for their success. Factors such as peer

coaching, teaming within and across grades, working together to help all students, and program consistency were mentioned as aspects of collaboration which teachers valued in these most effective schools. In contrast, in the moderately and least effective schools, concerns about program compatibility, instructional consistency, and common instructional terminology were more prevalent. One of the teachers at Wheeler, one of the most effective schools, summed it up this way:

> Teaming with other staff is important. You can't do it by yourself. Teaming also builds a sense of community. If the children see us working together and getting along, that means a lot to them. The children also get to see other teachers and get to know them. That builds caring and community.

This sentiment was echoed by teachers at Hilltop School, who suggested that peer coaching and collaboration, because they led to schoolwide buy-in, were key factors in their success.

Early reading interventions

Principals and teachers in three of the four most effective schools cited an externally developed, research-proven intervention as a part of their success. Woodlawn had small-group interventions in grades K–3 which met for approximately 30 minutes a day. Wheeler and Hilltop, K–2 schools, had daily small-group interventions in each grade. On the other hand, Stevenson had no externally developed early reading intervention but did have locally developed safety nets for struggling readers in the primary grades.

It is important to point out that the interventions in these most effective schools were not limited to grade 1 but spanned the primary grades. They were small-group interventions rather than one-on-one interventions. Also, three of the programs in the most effective schools had been developed through grassroots efforts within their regions, and the fourth was locally developed. None was a widely implemented, nationally disseminated reading intervention program or even a component of a nationally disseminated schoolwide reform program.

By comparison, two moderately effective schools operated a one-on-one nationally disseminated intervention for a relatively small number of students in grade 1. A third moderately effective school had a regionally developed one-on-one tutoring program in place for grade 1 students. A fourth moderately effective school had a regionally developed grade 1 tutoring program, and two had no intervention in place.

One of the least effective schools had a one-on-one intervention for grade 1 as part of a nationally recognized schoolwide reform program, and another had a small-group, regionally developed intervention for grade 3. Two of the least effective schools had no intervention in place. (See Table 1 for a summary of the types of interventions and school reform efforts in all schools.)

While the sample of schools in this study is limited, the data nonetheless point to reading success in the most effective schools through a combination of locally or regionally developed small-group interventions set within a homegrown school reform model. The experiences stand in contrast to a national push (e.g., the 17 programs listed in the Comprehensive School Reform Demonstration Program legislation; see Herman, 1999) for off-the-shelf reading intervention programs and school reform models (American

Federation of Teachers, 1997; Herman & Stringfield, 1997; Slavin & Fashola, 1998; Wang, Haertel, & Walberg, 1998).

Teachers in the most effective schools felt strongly that the early interventions in place in their buildings were key to their success. In the words of one Hilltop teacher, "Our early intervention makes all the difference in terms of taking care of kids' needs." One Woodlawn teacher puts it this way: "With our early reading intervention kids experience success and look forward to the time [in the intervention] ...because of the teacher and the success."

Ongoing professional development

During interviews, teachers and/or principals in three of the four most effective schools cited a yearlong staff development effort related to their early intervention program as responsible for their success, indicating that it helped them "stay in a learner mode," and "all be of one accord." For example, at Hilltop the teachers took two or three yearlong courses (at the beginner, experienced, or advanced level) on the philosophy of the intervention and the implementation of the strategies within the classroom. Teachers were encouraged to take at least two of the three courses, which met for two hours a month during the school year. There was also a time for every class participant to meet with a peer coach for 45 minutes once every two weeks.

In addition to the information about professional development related to the early reading interventions, teachers were asked in the survey about preferred approaches to professional development. In three of the most effective schools, a majority of teachers rated "visits to schools with innovative programs followed by sharing of observations with colleagues" as an effective approach to professional development that had been used in their school. Also, in three of the four most effective schools, a majority of teachers rated "district or school sponsored yearlong workshops" or "graduate-level courses" as an effective approach to professional development that had been used in their school. In two of the most effective schools, a majority of teachers rated "mentorship programs between experienced and new teachers" as effective. Approaches rated as ineffective by a majority of teachers in two of the most effective schools were "speakers and topics chosen by the district," and in three schools, "inservice provided by publishing companies with the adoption of a new textbook series."

All four of the most effective schools cited the importance of ongoing professional development when giving reasons for their success. In the words of a teacher from Hilltop, "You need to believe in change to be an educator, not be afraid to change, learn from experiences, and take classes to encourage growth."

Building collaboration

Building collaboration was positively and statistically significant related to fluency ($r = .43$) and retelling ($r = .35$). All four of the most effective schools had reorganized their instructional delivery system within the past few years to make use of a collaborative model for reading instruction. Typically, this meant that special personnel—a Title I, reading resource, or special education teacher—went into the classroom for an hour a day to help provide instruction for small, ability-based groups. Wheeler deployed a resource teacher in the classroom with the classroom teacher during reading time for one hour and an aide for a second hour of reading. Children received guidance in small groups with a teacher, a resource teacher or the aide in one-on-

one settings. At this school, a total of two and a half to three hours a day was spent on reading/language arts instruction and practice. Clearly, helping all children learn to read was a priority at this school. Woodlawn sent a reading specialist and a special education teacher into the classroom to work with small groups along with the classroom teacher for an hour each day. In this school, two and a half hours a day were spent on reading/language arts instruction in grades 1 and 2. At Stevenson, using a similar push-in collaborative model, children also received about two and a half hours a day of reading/language arts instruction and practice. A Title I teacher or aide worked in the classroom for 50 minutes each morning. The Title I teachers returned to the first- and second-grade classrooms for 30 minutes in the afternoon to provide one-on-one or one-on-two help to struggling readers. At Hilltop, teachers also used a collaborative model, but in this case the children who were struggling most in reading left the classroom during the two and a half to three hour literacy block to receive small-group instruction for 45 minutes. This small group instruction, delivered to two or three children at a time, was highly compatible with the instruction the children received in their regular classroom.

A common element of all four of these buildingwide approaches was the focus on small-group instruction. In three instances, special teachers came into the regular classrooms to help provide this small-group instruction, while in a fourth, children left to receive small-group intervention in a workshop setting. In all four of these most effective schools, teachers spent a large amount of time, averaging 135 minutes a day, on reading instruction. In interviews, teachers in three of these schools mentioned that reading was a priority at their school; their time allocation to reading is strong evidence of this commitment.

Additionally, as reported above, three or the four schools had recently implemented a regionally developed, research-based reading intervention program from kindergarten through grade 2 or 3. These efforts had been tied to buildingwide ongoing professional development spanning more than a year. Additionally, these efforts undoubtedly contributed to positive communication of ideas across teachers, which Wilder (1977) had identified as important in most of the effective high-poverty schools he studied. In fact, the simple correlations at the building level between building communication and other features of effective schools was consistently positive: $r = .79$ with the presence of an internal assessment system, $r = .64$ with parent links, and $r = .40$ (*ns*) with the presence of an intervention. The findings are consistent with the findings of Stringfield et al. (1997) that among schools implementing special strategies, those showing the greatest achievement gains had worked hard at initial and long-term maintenance of their innovations. In words of advice to other schools wishing to significantly increase their reading achievement, the principal at Wheeler offered the following: "Identify what you want to do and work collaboratively. Embrace one thing and stick to it. Also, look at what's working and what's not and make adaptations through continuous progress monitoring."

Summary

Across the four most effective schools in this study, reading was clearly a priority. The teachers and principals considered reading instruction their job and they worked at it. They worked together, worked with parents, and worked with a positive attitude to reach the goal of all children reading well before they left the primary grades. They set personal preferences aside in

order to reach consensus on schoolwide monitoring systems, curriculum, and professional development, with the constant goal of improving an already effective reading program.

Instructional Practices Within Levels of School Effectiveness

Multivariate analysis

To investigate the relationship between school effectiveness and classroom instruction, we initially conducted a multivariate analysis of variance (MANOVA) with the school effectiveness rating serving as the independent variable and eight teacher variables serving as outcome measures (see Table 11). To ensure that we were focusing on potentially powerful variables, only

Table 11: ANOVAs on Teacher Factors With School Effectiveness Rating

	F	*P*	School Rating*	Mean	*SD*	*N*†
Home communication (mean is based on 5-level rating)	5.25	.008	1 2 3 3 ≠ 2 = 1	3.09 2.95 3.75	1.22 1.12 1.03	22 27 19
Engagement (mean is based on 3-level rating)	.82	*ns*	1 2 3	2.18 2.24 2.40	.73 .74 .84	22 29 19
Time in small-group instruction (mean is minutes/day)	9.63	.0002	1 2 3 3 ≠ 2 = 1	37.94 26.00 59.82	27.19 20.16 28.46	20 24 19
Time in whole-group instruction (mean is minutes/day)	2.20	*ns*	1 2 3	30.20 36.66 24.86	18.42 20.93 14.64	20 24 19
Time in independent reading (mean is minutes/day)	4.24	.02	1 2 3 3 = 2 ≠ 1	18.63 27.04 28.14	9.43 13.30 10.50	20 24 19
Preferred interaction style of coaching	2.32	*ns*	1 2 3	.23 .21 .47	.43 .41 .51	22 29 19
Preferred interaction style of telling	2.01	*ns*	1 2 3	.41 .41 .16	.50 .50 .37	22 29 19
Preferred interaction style of recitation	.17	*ns*	1 2 3	.32 .34 .26	.48 .48 .45	22 29 19

* 1 = Least effective; 2 = moderately effective; 3 = most effective.

† Number of teachers.

those classroom factors which were statistically significantly related to one or more of the measures of student or teacher accomplishment (school effectiveness rating; fluency, retelling, or reading words measure; or teacher

accomplishment rating) were included in the MANOVA. A statistically significant MANOVA, $F(14, 108) = 2.56$, $p < .01$, led us to conduct follow-up univariate analyses of variance (ANOVAs; see Table 11).

Univariate analysis

Home communication. The follow-up ANOVA on home communication was statistically significant, $F(2, 65) = 5.25$, $p < .01$. Tukey post hoc tests revealed that the teachers in the most effective schools communicated more with parents/caretakers than teachers in the moderately effective or least effective schools.

Teachers in the most effective schools were more likely to send a letter or newsletter home weekly and call home regularly than teachers in the other schools. Overall, a higher proportion of teachers (15/23) in the most effective schools rated high on the home linkage scale than in either the moderately (8/29) or least (10/29) effective schools. In three of the four most effective schools, more than half of the teachers reported calling home at least once a month. In only one of six moderately effective schools and one of four of the least effective schools did a majority of teachers indicate that they called home as frequently. In one of the most effective schools, teachers made a concerted effort to call home with positive comments, and 51% of parents who were asked said they had received such calls. Looking at both teacher and school factors, the analysis of links to families suggests that personnel in the most effective schools made a more concerted effort than personnel in other schools to reach out to parents.

Student engagement. The univariate ANOVA revealed no statistically significant school effectiveness effect for student engagement rating, $F(2, 67) = .82$, $p > .05$, across levels of school effectiveness. As we will see later, this result is to be contrasted with the finding for levels of teacher accomplishment.

Time spent in small- and whole-group instruction. The ANOVA on time in small-group instruction revealed a statistically significant effect for level of school effectiveness, $F(2, 60) = 9.63$, $p < .001$, but not for grade or the interaction between the two. Tukey post hoc tests revealed that students of teachers in the most effective schools spent more time daily in small-group instruction ($M = 59.02$ minutes per day) than students of teachers in the moderately effective schools ($M = 26.10$ mpd) or least effective schools ($M = 37.94$ mpd). The one-way ANOVA for school effectiveness rating on time spent in whole-group instruction was not statistically significant, $F(2, 60) = 2.20$, $p > .05$. However, even the students in the most effective schools, who were spending an hour a day in small groups, were in whole-class instructional activities across an average of 25 minutes per day.

In addition to differences in the amount of time spent in small-group instruction by school effectiveness, the ratio of small- to whole-group instruction is important to consider. In each of grades 1, 2, and 3, children in the most effective schools spent more time in small-group instruction than in whole-group instruction. In grades 1 and 2, the small group/large group ratio was 2/1. The ratio was this high only in grade 1 in the least effective schools. In fact, for children in grade 3 in the moderately effective and least effective schools the ratio was 1/2 (see Table 12).

Table 12: Mean Time in Instruction (and Standard Deviation) by Grade Level and School Rating

	School Rating*	N†	1 Time in Whole Group	2 Time in Small Group	3 Time in Independent Activity (Excldg #s 4 & 5)	4 Independent Reading	5 Writing in Response to Reading	6 Total Time in Reading
Grade 1	1	7	19.27 (20.39)	47.30 (22.09)	4.69 (7.78)	15.30 (10.08)	18.07 (11.97)	104.16 (21.73)
	2	7	29.06 (17.78)	28.56 (22.20)	12.43 (20.74)	28.66 (14.92)	17.14 (13.46)	115.84 (49.20)
	3	8	23.05 (15.51)	66.96 (38.50)	11.68 (12.54)	27.79 (7.04)	17.44 (11.13)	146.91 (35.46)
	Grade mean	22	23.76 (17.48)	48.49 (32.21)	9.69 (14.32)	24.09 (12.13)	17.55 (11.58)	123.57 (39.91)
Grade 2	1	8	29.05 (8.91)	38.23 (32.16)	13.15 (16.95)	18.11 (9.36)	19.65 (7.72)	118.19 (20.65)
	2	7	30.54 (22.86)	37.20 (22.59)	11.33 (7.82)	21.99 (11.85)	9.21 (9.91)	110.27 (34.79)
	3	7	23.93 (8.58)	58.50 (21.68)	4.17 (4.62)	29.66 (12.11)	13.19 (6.47)	139.49 (25.80)
	Grade mean	22	27.90 (14.37)	44.35 (26.89)	9.71 (11.67)	23.02 (11.64)	14.25 (8.93)	119.25 (27.26)
Grade 3	1	5	47.34 (16.41)	24.38 (24.23)	8.14 (8.31)	24.10 (7.67)	13.98 (13.27)	117.94 (19.70)
	2	10	46.27 (19.58)	16.61 (13.02)	3.37 (4.74)	29.44 (13.47)	16.40 (11.79)	112.09 (33.88)
	3	4	30.15 (22.99)	47.85 (11.42)	3.75 (7.50)	26.20 (15.55)	9.70 (4.89)	117.65 (46.76)
	Grade mean	19	43.16 (19.68)	25.23 (19.80)	4.71 (6.36)	27.35 (12.24)	14.35 (10.94)	114.80 (32.08)
School rating‡	1	20	30.20 (18.42)	37.94 (27.19)	8.94 (12.40)	18.63 (9.43)	17.68 (10.47)	113.38 (20.68)
	2	24	36.66 (20.93)	26.10 (20.16)	8.33 (12.47)	27.04 (13.30)	14.52 (11.81)	112.66 (37.44)
	3	19	24.86 (14.64)	59.82 (28.46)	7.24 (9.63)	28.14 (10.50)	14.24 (8.70)	134.34 (34.88)
Mean across teachers			31.05 (18.78)	40.03 (28.40)	8.20 (11.50)	24.70 (11.94)	15.44 (10.47)	119.42 (33.19)

* 1 = Least effective; 2 = moderately effective; 3 = most effective.

† Number of teachers.

‡ Pooled across grades.

When asked on the survey to select the four most important factors for improving struggling readers' achievement, 83% of the teachers in the four most effective schools selected small-group instruction as an important factor. Additionally, in two of the schools, teachers mentioned the focus on small-group instruction as a factor contributing to their success. "Small, flexible groups at students' instructional level are important. They need to be coached at their instructional level" contended a Hilltop teacher. Comments about the virtues of small-group instruction were commonly heard in the

interviews at the most effective schools, such as those by a Woodlawn teacher—"Small groups or one-on-one every day really makes a difference"—and a Wheeler teacher—"Small groups are crucial. Children are more likely to succeed when they are in two groups of 6 with two teachers than when there are 12 children with one teacher."

In all of the most effective schools, the basis for forming the small groups was perceived ability, an observation which suggests that these teachers were more concerned about meeting students at their instructional level than they were about any damage to children's self-worth that might accrue from being a part of a group socially and personally sanctioned as the low group. Even so, the teachers in the most effective schools were very aware of the need to make sure that the groups were flexible, that students moved to another group when their performance (as measured by their internal school-based monitoring system) merited movement. The importance of schoolwide monitoring cannot be underestimated in this regard. These data provided teachers with regular, recurring opportunities to reflect upon the validity of their instructional groupings and modify membership accordingly. The principal at Stevenson talked about a mental change which occurred when her school became a schoolwide Title I building. Teachers began to talk more about children's needs during the year and made changes in students' reading group placement. "A reading group was no longer a year-long placement for children, particularly for the lowest children." At Woodlawn, teachers also voiced commitment to the idea of flexible ability groups. They used running records and periodic measures of words read correctly in one minute to move children to a higher group whenever they could. Furthermore, in three of the four schools, early reading interventions were in place across the primary grades to provide high-quality, special assistance to children who were struggling to learn to read. Recall also that in these most effective schools, students also averaged 25 minutes a day of whole-group instruction in which they were interacting across ability levels.

In short, the ability grouping in these schools was not a lifetime sentence to low group membership so powerfully documented in the literature on grouping; to the contrary, some of the special grouping practices, namely the special, supplemental instruction, were in place to accelerate struggling readers' literacy learning to the point where they could re-enter regular classroom groupings.

Time spent in independent reading. The ANOVA on time spent in independent reading was statistically significant, $F(2, 60) = 4.24$, $p < .05$. Tukey post hoc tests revealed that students in the most effective schools (M = 28.14 minutes per day) and moderately effective schools (M = 27.04 mpd) spent more time in independent reading than students in the least effective schools (M = 18.63 mpd).

In three of the most effective schools, teachers mentioned providing time for students to read authentic texts as a factor contributing to their school's success. "I give my students lots of time to engage in reading/writing opportunities. Lots of opportunities to read all kinds of texts," explained a Hilltop teacher. Or, as a Woodlawn teacher put it, "You become a better reader by reading. My students read at least 20–30 minutes a day. Also, partner reading—they love it." "Everyone in the whole school is taking books home at night for reading. It's one of our school improvement goals," pointed out

another Hilltop teacher. These findings complement earlier research documenting that time spent in independent reading in school does make a difference in students' reading achievement (Anderson, Wilson, & Fielding, 1988; Elley & Mangubhi, 1983; Taylor, Frye, & Maruyama, 1990).

Preferred interaction style. While the trends were provocative, with half the teachers in the most effective schools preferring coaching compared to about a quarter of teachers in the moderately and least effective schools, the ANOVAs on preferred interaction styles by school effectiveness level were not statistically significant—coaching, $F(2, 67) = 2.32$, $p > .05$; telling, $F(2, 67) = 2.01$, $p > .05$; or recitation, $F(2, 67) = .17$ (see Table 11).

Supplementary analyses of reading specific teaching strategies

In addition to the question of the use and impact of more generic teaching styles, we were able to apply nonparametric analyses to two additional reading-specific teaching domains—word recognition and comprehension instruction. In the case of word recognition, we were limited to grades 1 and 2 because of the paucity of word recognition instruction observed in grade 3.

Approach to word recognition. Chi square tests revealed that in comparison to the moderately effective schools ($\chi^2 = 5.0$) and least effective schools ($\chi^2 = 5.4$), more grade 1 and 2 teachers in the most effective schools were frequently observed coaching as children were reading to teach word recognition. More teachers in the most effective ($\chi^2 = 5.5$), and least effective schools ($\chi^2 = 8.5$), were frequently observed practicing sight words than teachers in the moderately effective schools. There were no differences in the incidence of teachers in grades 1 and 2 who provided explicit phonics instruction across the three types of schools (see Table 13).

Table 13: Approaches to Word Recognition Instruction: Frequency of Teachers in Grades 1 and 2 (by School Effectiveness Rating) Observed Using a Particular Strategy

		Coaching in Word Recognition Strategies During Reading	Providing Explicit Phonics Instruction	Practicing Sight Words
School rating*	1	2/15 (13%)	9/15 (60%)	6/15 (49%)
	2	3/18 (17%)	11/18 (61%)	0/18 (0%)
	3	8/15 (53%) $3 > 2, \chi^2 = 5.0, p < .05$ $3 > 1, \chi^2 = 5.4, p < .05$	9/15 (60%)	4/15 (27%) $3 > 2, \chi^2 = 5.5, p < .05$ $1 > 2, \chi^2 = 8.5, p < .05$

* 1 = Least effective; 2 = moderately effective; 3 = most effective.

In third grade, very little word recognition instruction was observed anywhere. In one most effective school, third-grade teachers provided small-group instruction to struggling readers in which they coached them to decode multisyllabic words as they were reading. In another most effective school and in one moderately effective school, struggling third-grade readers

received small-group instruction from resource teachers who used a combination of coaching while reading and work on word families to teach word recognition. In one most effective school and three moderately effective schools, third graders continued to do word study as a subject separate from reading.

While not universal across all teachers, there is a definite trend, in the most effective schools, for grade 1 and 2 teachers to combine (a) explicit phonics instruction in isolation with (b) coaching students to use a range of strategies to figure out unknown words when they encounter them in everyday reading. In contrast, the teachers in the moderately effective schools primarily provided phonics instruction, with only a few adding the coaching component. In the least effective schools, teachers primarily provided explicit phonics instruction, with about half adding practice on sight words.

When asked about reasons for their success, teachers in two of the most effective schools mentioned the importance of teaching students strategies, not skills, as a factor contributing to their success. As one Hilltop teacher put it, "I focus on strategies rather than specific skills—metacognitive strategies, demonstrations, how to do think-alouds. I am process oriented so kids become independent rather than reliant on the teacher." This sentiment was echoed by a Woodside teacher: "You're not going to improve as a teacher if you don't get into teaching strategies. I teach my students strategies to become independent."

Approach to comprehension instruction. Chi square tests revealed that there were more teachers in the most effective schools frequently observed asking higher level questions about stories students had read than teachers in moderately effective (2 of 29, $\chi^2 = 9.1$, $p < .01$) or least effective (0 of 22, $\chi^2 = 9.8$, $p < .01$) schools (see Table 14). That said, we must reiterate the extremely low rate of these more cognitively challenging activities in the overall sample.

Table 14: Approaches to Comprehension Instruction: Frequency of Teachers (by School Effectiveness Rating) Observed Using a Particular Strategy

		Asking Text-Based Questions	Asking Aesthetic Response Questions	Having Students Write in Response to Reading
School rating*	1	10/22 (45%)	0/22 (0%)	6/22 (27%)
	2	10/29 (34%)	2/29 (7%)	7/29 (24%)
	3	7/19 (37%)	7/19 (37%) 3 > 2, $\chi^2 = 9.1$, $p < .01$ 3 > 1, $\chi^2 = 9.8$, $p < .01$	9/19 (47%)

* 1 = Least effective; 2 = moderately effective; 3 = most effective.

Word recognition work and reading practice were much more the focus of reading instruction in grades 1-2 across all schools in this study than was

comprehension. Explicit instruction in comprehension strategies was seldom witnessed across grades 1-3. Discussions which stretched children's thinking were also infrequent across grades 1-3.

The Relationship of Instructional Practice to Teacher Accomplishment

The analyses of instructional practices within levels of school effectiveness document the fact that, on average, teachers within effective schools operate differently than do teachers in other schools. These average differences, however, mask instructional variation among teachers within schools. Not all of the best teachers worked in the most effective schools. In order to gain

Table 15: Correlations for Teacher Rating and Teacher Factors

	School Rating	Teacher Rating	Z-score Residual Fluency (Class Mean)	Z-score Residual Retelling (Class Mean)	Z-score Residual Words (Class Mean)
Teacher rating	.23 $p = .05$ $n = 71$		.27 $p < .001$ $n = 69$	.10 *ns* $n = 69$	.18 *ns* $n = 69$
Home communication	.26 $p = .03$ $n = 68$	.27 $p = .02$ $n = 68$	.06 *ns* $n = 66$	.25 $p = .04$ $n = 66$	-.09 *ns* $n = 66$
Time on task	.14 *ns* $n = 70$	.79 $p < .001$ $n = 70$	.21 *ns* $n = 68$	.17 *ns* $n = 68$	.18 *ns* $n = 68$
Time spent in small-group instruction	.30 $p = .02$ $n = 63$	.31 $p = .02$ $n = 63$	.41 $p < .01$ $n = 61$	.10 *ns* $n = 61$	.22 *ns* $n = 61$
Time spent in whole-group instruction	-.11 *ns* $n = 63$	-.41 $p < .01$ $n = 63$	-.03 *ns* $n = 61$	-.22 *ns* $n = 61$	.06 *ns* $n = 61$
Time spent in independent reading	.32 $p = .01$ $n = 63$	.10 *ns* $n = 63$	.33 $p < .01$ $n = 61$	-.11 *ns* $n = 61$	.20 *ns* $n = 61$
Preferred teaching style: coaching/ scaffolding	.20 *ns* $n = 70$	.40 $p < .01$ $n = 70$	.07 *ns* $n = 68$	-.18 *ns* $n = 68$	-.03 *ns* $n = 68$
Preferred teaching style: telling information	-.20 *ns* $n = 70$	-.53 $p < .01$ $n = 70$	-.27 $p = .03$ $n = 70$	-.14 *ns* $n = 70$	-.01 *ns* $n = 70$
Preferred teaching style: recitation	-.04 *ns* $n = 70$	.04 *ns* $n = 70$	.18 *ns* $n = 68$	.24 $p = .05$ $n = 69$	.07 *ns* $n = 68$
Time spent writing in response to reading	-.13 *ns* $n = 63$	.03 *ns* $n = 63$	-.18 *ns* $n = 61$	.03 *ns* $n = 61$	.01 *ns* $n = 61$
Time spent reading aloud	.17 *ns* $n = 63$	.14 *ns* $n = 63$	.19 *ns* $n = 61$	-.07 *ns* $n = 61$	.14 *ns* $n = 61$

additional insight into other factors that might lead to explanations of how to nurture teaching strategies that promote student learning, we undertook an analysis of instructional practices that was independent of student achievement. Instead, we used the ratings assigned to teachers on a joint (effective teaching and culturally responsive pedagogy) teacher accomplishment scale to classify teachers into three levels of accomplishment (most, moderately, and least); these levels were used as predictor variables to explain variations in the instructional practices used by teachers (Table 15).

Table 16: ANOVAs on Teacher Factors With Teacher Rating

	F	*P*	Teacher Rating	Mean	*SD*	*N**
Home communication (mean is based on 5-level rating)	2.4	*ns*	1 2 3	2.88 3.23 3.64	1.22 1.23 1.03	17 23 28
Time on task (mean is based on 3-level rating)	85.41	< .001	1 2 3 3 ≠ 2 ≠ 1	1.29 2.21 2.93	.47 .51 .26	17 23 29
Time in small-group instruction (mean is minutes/day)	3.08	.05	1 2 3 3 ≠ 1	25.35 38.67 48.25	26.09 33.00 21.51	13 23 27
Time in whole-group instruction (mean is minutes/day)	8.66	< .01	1 2 3 3 = 2 ≠ 1	47.94 28.98 24.69	18.38 17.12 15.78	13 23 27
Time in independent reading (mean is minutes/day)	.87	*ns*	1 2 3	23.97 22.51 26.92	11.77 10.98 12.81	13 23 27
Preferred interaction style of coaching	5.92	< .01	1 2 3 3 ≠ 2 = 1	.06 .21 .48	.24 .41 .51	17 24 29
Preferred interaction style of telling	16.60	< .001	1 2 3 3 ≠ 2 ≠ 1	.745 .375 .069	.440 .491 .258	17 24 29
Preferred interaction style of recitation	1.00	*ns*	1 2 3	.18 .38 .34	.39 .49 .48	17 24 29

* Number of teachers.

To investigate the relationships among the various indicators of teacher expertise and classroom practices, we subjected this large set of teacher variables to a MANOVA. We used three levels of the teacher accomplishment rating (most, moderately, and least accomplished) as the independent variable and eight scores from the teacher factor set derived from our empirical data (time spent in small-group instruction, time spent in whole-group

instruction, time spent in independent reading, student engagement rating, home communication rating, preferred style of telling, preferred style of recitation, and preferred style of coaching) as the set of dependent measures. The MANOVA was statistically significant, $F(14, 108) = 10.77, p < .001$, so follow-up univariate ANOVAs were conducted (see Table 16).

Communicating to parents

The ANOVA on level of home communication was not statistically significant, $F(2, 65) = 2.40, p > .05$. That is, no differences were observed in the frequency with which teachers of different levels of accomplishment communicated with students' parents or guardians. Recall that this factor was statistically significant in both school-level analyses; that is, most effective schools had higher composite school ratings and teachers in the most effective schools had higher ratings on the same home communication scale. The differences between the school level and the accomplishment analyses suggest either that the most accomplished teachers are not necessarily the best communicators or that teacher effects are moderated by a school-level ethic for this type of activity.

Student engagement

The ANOVA for the student engagement rating was statistically significant, $F(2, 67) = 85.41, p < .001$. Tukey post hoc tests revealed that the most accomplished teachers had a higher rating for keeping students engaged ($M = 2.93$ out of a possible 3) than the moderately accomplished teachers ($M = 2.21$), who, in turn, had a higher mean score than the least accomplished teachers ($M = 1.31$).

To shed more light on this teacher factor, in six of our sites, we were able to conduct a special analysis to learn even more about the impact of the student engagement variable. During the last two observations, observers were asked to interrupt their normal observational protocol every five minutes, scan the room quickly, and record the proportion of children in the class who were perceived to be on task. Grade 1–3 teachers rated as most accomplished were found to have an average of 96% of their students on task when the five-minute counts of students on task were taken. By contrast, the engagement rates were 84% and 61%, respectively, for the moderately accomplished and least accomplished teachers. Since these numbers are based on only 30 teachers and 60 observations, they should be interpreted cautiously. Even so, they underscore the importance of student engagement as a key curricular and management concern for teachers as they implement reading programs in their classrooms. The findings suggest that, unlike parent communication, wherein individual teacher practices appear to be moderated by school-level efforts, promoting high student engagement is a teaching practice not easily influenced by school-level practice. Developing the disposition and the curriculum, instruction, and interaction tools required to engage students may require more time, mentoring, and reflection than can be accomplished in typical inservice sessions.

Grouping practices

The ANOVA on time spent in whole group, $F(2, 60) = 8.66, p < .01$, indicated that students with teachers rated as least accomplished spent more time in whole-group instruction ($M = 47.94$ minutes per day) than teachers rated as moderately accomplished ($M = 28.98$ mpd) or teachers rated as most accomplished ($M = 24.69$ mpd). A two-way (teacher accomplishment by grade) ANOVA showed a statistically significant effect for grade level, $F(2, 54) = 7.90, p < .01$, with a strong tendency for whole-group time allocations to

increase with grade level, but no statistically significant grade by teacher accomplishment interaction. Means and standard deviations by grade level and teacher rating are shown in Table 17.

Table 17: Mean Time in Instruction (and Standard Deviation) by Grade Level and Teacher Rating

	Teacher Rating*	N†	1 Time in Whole Group	2 Time in Small Group	3 Time in Independent Activity (Excldg #s 4 & 5)	4 Independent Reading	5 Writing in Response to Reading	6 Total Time in Reading
Grade 1	1	3	50.00 (3.06)	32.83 (35.92)	15.40 (5.70)	15.97 (10.06)	25.70 (11.29)	123.70 (38.78)
	2	7	23.53 (15.71)	49.17 (47.98)	10.34 (4.23)	23.44 (13.28)	9.87 (7.61)	116.36 (48.02)
	3	12	17.33 (14.64)	52.00 (20.24)	7.88 (11.27)	26.50 (11.90)	19.98 (11.80)	139.90 (30.90)
	Grade mean	22	23.76 (17.48)	48.49 (32.21)	9.69 (14.32)	24.09 (12.13)	17.55 (11.58)	123.57 (39.91)
Grade 2	1	6	36.45 (18.09)	28.25 (27.46)	9.05 (10.67)	20.95 (10.75)	14.25 (10.09)	108.95 (22.17)
	2	7	26.79 (13.12)	47.34 (29.00)	45.60 (16.55)	21.19 (11.42)	16.76 (7.58)	127.14 (24.67)
	3	9	23.06 (11.05)	52.78 (22.66)	5.58 (5.26)	25.82 (13.06)	12.74 (9.83)	119.98 (32.37)
	Grade mean	22	27.90 (14.37)	44.35 (26.89)	9.71 (11.67)	23.02 (11.64)	14.25 (8.93)	119.25 (27.26)
Grade 3	1	4	63.63 (14.25)	15.38 (20.05)	10.03 (8.84)	34.50 (7.98)	16.80 (8.52)	140.33 (26.60)
	2	9	34.93 (20.53)	23.79 (19.85)	3.30 (4.86)	22.81 (9.95)	9.81 (5.08)	94.65 (23.47)
	3	6	41.85 (11.80)	33.97 (19.16)	3.27 (5.69)	25.40 (16.11)	19.53 (16.50)	128.02 (30.62)
	Grade mean	19	43.16 (19.68)	25.23 (19.80)	4.71 (6.36)	27.35 (12.24)	14.35 (10.94)	114.80 (32.08)
Teacher rating‡	1	13	47.94 (18.38)	25.35 (26.05)	10.82 (8.91)	23.97 (11.77)	17.68 (10.20)	125.75 (28.33)
	2	23	28.98 (17.12)	38.67 (33.91)	9.15 (15.30)	22.51 (10.98)	11.79 (7.07)	111.14 (34.59)
	3	27	24.65 (15.78)	48.25 (21.51)	6.09 (8.48)	26.92 (12.81)	17.47 (17.35)	123.42 (33.87)
Mean across teachers		63	31.05 (18.78)	40.03 (28.40)	8.20 (11.50)	24.70 (11.94)	15.44 (10.47)	119.42 (33.19)

* 1 = Least accomplished; 2 = moderately accomplished; 3 = most accomplished.

† Number of teachers.

‡ Pooled across grades.

The ANOVA on time spent in small-group instruction revealed an effect for level of teacher accomplishment, $F(2, 60) = 3.08$, $p = .05$, with students in the classrooms of teachers rated as most accomplished spending more time in small-group instruction ($M = 48.25$ minutes per day) than students with

teachers rated as moderately accomplished (M = 38.67 mpd), who, in turn, spent more time than students with teachers rated as least accomplished (M =25.35 mpd).

The ANOVA on time spent on independent reading indicated no statistically significant differences between teachers with different levels of ratings on the accomplishment scale. Students averaged from 23 to 27 minutes a day in independent reading across all conditions of teacher effectiveness. This finding is at variance with the parallel analysis for teachers in the most effective schools, where reliable school effectiveness differences emerged for independent reading. As with the parent outreach finding, it suggests that teacher practices for independent reading may have been moderated by school level initiatives and/or philosophies.

Interaction patterns

The ANOVA on preferred interaction style of coaching and teacher accomplishment was statistically significant, $F(2, 67) = 5.92$, $p < .01$. Tukey post hoc tests revealed that more of the most accomplished teachers had a preferred interaction style of coaching than the moderately or least accomplished teachers (see Table 18 for a complete presentation of the interaction preferences). A higher incidence of most accomplished teachers exhibited a preferred interaction style of coaching (n = 13; 48%) than did the least accomplished teachers (n = 1; 6%) or moderately accomplished teachers (n = 5; 21%).

Table 18: Relationship Between Teacher Accomplishment and Preferred Style of Interaction

	Teacher Rating		
	Least Accomplished	Moderately Accomplished	Most Accomplished
Preference for coaching			
Yes	1 (6%)	5 (21%)	14 (48%)
No	16	19	15
Total	17	24	29
Preference for telling			
Yes	13 (75%)	9 (38%)	2 (7%)
No	4	15	27
Total	17	24	29
Preference for recitation			
Yes	3 (18%)	9 (38%)	10 (34%)
No	14	15	19
Total	17	24	29

The ANOVA on preferred interaction style of telling was statistically significant, $F(2, 67) = 16.60$, $p < .001$. Tukey post hoc tests revealed that more least accomplished teachers (n =13; 75%) preferred telling than moderately accomplished teachers (n = 9; 38%). There were also more moderately accomplished teachers with a preferred interaction style of telling than teachers rated as most accomplished (n = 3; 7%). The ANOVA on preferred interaction style of recitation and teacher accomplishment was not statistically significant. The data on coaching and telling may well be two sides of the same coin. Even though our observation system allowed for the possibility that a teacher could receive both codes within a given five-minute obser-

vational block, that is not what we found. For individual teachers, the one seems to be an alternative to another, and preferred styles can be predicted by teacher accomplishment.

In contrast to statistically nonsignificant differences for the teachers within levels of school effectiveness, these statistically significant differences among teachers across schools suggest that a teacher's preferred style of interacting with students is a teaching dimension which is less well influenced by the practice of others at the school level than other dimensions of teaching being investigated in our study (e.g., time spent by students in independent reading, or degree of home communication). As with high levels of student engagement, a preferred style of coaching during reading instruction may be a teaching skill which requires time and/or support from more accomplished teachers to develop.

Approaches to word recognition instruction

Of the 22 teachers in grades 1 and 2 who were rated high on the composite teacher accomplishment rating (effective teaching and culturally responsive teaching), 10 (45%) were frequently observed coaching children on how to use different word recognition strategies to figure out unknown words while they were reading connected text. By contrast, of the 15 moderately accomplished teachers in grades 1 and 2, 3 (20%) were observed frequently coaching children as they were reading. Of the 11 teachers perceived as least accomplished, none was frequently observed using the coaching-while-reading strategy to teach word recognition. Chi square tests confirmed the fact that these differences were statistically significant.

In contrast, no differences were seen across teacher effectiveness ratings in terms of providing explicit phonics instruction. Thirteen of 22 most accomplished teachers, 11 of 15 moderately accomplished teachers, and 5 of 11 least accomplished teachers were frequently observed providing explicit phonics instruction (see Table 19). Chi square tests revealed no statistically significant differences according to teacher accomplishment rating in the frequency of teachers frequently observed providing explicit phonics instruction.

Table 19: Approaches to Word Recognition Instruction: Frequency of Teachers in Grades 1 or 2 (by Teacher Accomplishment Rating) Using a Particular Strategy

		Coaching on Word Recognition Strategies During Reading	Providing Explicit Phonics Instruction	Practicing Sight Words
Teacher accomplishment rating*	1	0/11 (0%)	5/11 (45%)	1/11 (9%)
	2	3/5 (20%)	11/15 (73%)	1/15 (7%)
	3	10/22 (45%) $3 > 1, \chi^2 = 7.17, p < .05$	13/22 (59%)	8/22 (36%) $3 > 2, \chi^2 = 4.27, p < .05$

* 1 = Least accomplished; 2 = moderately accomplished; 3 = most accomplished.

Eight of the 22 most accomplished teachers were frequently observed reviewing sight words, in contrast to 1 of 15 moderately accomplished and 1

of 11 least accomplished teachers. A chi square test indicated that more teachers identified as most accomplished were frequently observed practicing sight words than teachers identified as moderately accomplished.

What emerges is a pattern in which the most accomplished teachers demonstrate a more balanced portfolio of approaches to assist in word identification (i.e., more of them do a little of each practice) and are, by and large, the only group of teachers who go the extra mile in helping students apply the alphabetic principle to work in everyday reading tasks.

Comprehension instruction

Across all schools, comprehension instruction was minimal in grades 1-3. Primary modes of working on comprehension included asking questions (many of which were literal) about the story as children were reading, either in small groups or in a whole-class setting, and having children write in response to stories they had read. This writing was most typically in the form of a journal entry or answers to written questions. Twenty-nine of 70 teachers were frequently observed asking text-based questions, and 27 were frequently observed having children write in response to what they had read. Only 11 of 70 teachers were seen frequently asking higher level questions about children's feelings or about their lives in relation to a story they had read. Only 5 teachers were frequently observed providing instruction (not including worksheet completion) about a comprehension skill or strategy (see Table 20.)

Table 20: Number of Grade 1–3 Teachers (out of 68) Observed Using a Particular Comprehension Technique

	Frequently Observed	Occasionally Observed
Doing a picture walk	5	9
Asking for a prediction	6	20
Asking text-based questions	29	22
Asking aesthetic response questions	11	16
Having children write in response to reading	27	26
Doing a story map	2	16
Having children retell a story	2	20
Working on a comprehension skill or strategy	5	22

Nonetheless, when looking at teachers across the three levels of teacher accomplishment, chi square tests revealed that these practices were not randomly distributed. Compared to least accomplished teachers, more most accomplished teachers were frequently observed asking higher level questions (9 of 29 versus 0 of 17) and having students write in response to reading (14 of 29 versus 3 of 17). Also, more (9 of 29) of the most accomplished teachers were frequently observed asking higher level questions that the moderately accomplished teachers (2 of 24). Chi square tests for text-based questions were not statistically significant (see Table 21). In the bigger picture, however, what we found in comprehension instruction is disconcerting. The low overall incidence of teachers in the entire sample asking higher level questions should concern us: Differences among schools notwithstanding, only 16% of the teachers in grades 1-3 in this study were frequently observed asking higher level, aesthetic response questions. Little or no dis-

cussion was the dominant finding, followed closely by discussion focusing on the facts of the selection.

Table 21: Approaches to Comprehension Instruction: Frequency of Teachers (by Teacher Accomplishment Rating) Observed Using a Particular Strategy

		Text-Based Oral Questions	Higher Level (Aesthetic Response) Oral Questions	Students Write in Response to Reading
Teacher accomplishment rating*	1	4/17 (24%)	0/17 (0%)	3/17 (18%)
	2	11/24 (46%)	2/24 (8%)	10/24 (42%)
	3	14/29 (48%)	9/29 (31%) $3 > 1, \chi^2 = 6.56, p < .05$ $3 > 2, \chi^2 = 4.11, p < .05$	14/29 (48%) $3 > 1, \chi^2 = 4.32, p < .05$

* 1 = Least accomplished; 2 = moderately accomplished; 3 = most accomplished.

Looking Across Schools and Classrooms

For our final analysis, we conducted a stepwise regression in which the most powerful school level (systematic internal assessment and parent links) and classroom level (time in small-group instruction and time in independent reading) variables were simultaneously regressed on our most robust outcome measure, fluency as indexed by words correct per minute on a grade level passage. Parent links entered first with an R^2 of .35; time spent in small group instruction entered second, adding .09 to the R^2 for a total R^2 of .44. Systematic assessment and independent reading did not enter the equation, most likely due to strong covariation with one of the other factors that did enter.

As we learned from the descriptive analyses, the 60 minutes a day of small-group instruction were made possible because of collaborative efforts and negotiations among staff members working in a collaborative model. Also, schoolwide efforts and individual teacher efforts were considered in the parent links rating. It is important to highlight these programmatic effects because they remind us that school-level change is as important as change within classrooms.

Overall Discussion

Given the complexity of this research, we have tried to provide elaborate discussions of findings as they were presented. The primary role of the overall discussion is to consider relations among the various findings. We begin

with synthesis of key findings, trying to emphasize connections among the variables. Then we move to a more detailed discussion of the ways in which school-level and classroom-level factors interact to create different instructional environments. We end by contextualizing our work within the long-standing effective schools and effective teaching traditions, confessing our limitations, and providing a few broad conclusions.

Revisiting Key Findings

In this section, we look across all four of the previous analyses (school-level analyses of reading program characteristics, the practices of teachers within levels of school effectiveness, the practices of accomplished teachers independent of schools, and relationships among variables across schools and classrooms). One contribution of the current study is that it focused on classrooms as well as schools to get a richer picture of what was happening in schools that excel at promoting growth among struggling readers.

Management and engagement

As has been found in the research on effective teachers (Brophy, 1973; Wharton-MacDonald et al., 1998), the most accomplished teachers in this study were experts at classroom management, as reflected in the summaries of observations and in the time-on-task ratings. In general, they had well-established classroom routines and procedures for handing behavior problems, quick transitions between activities, and a rapid pace of instruction, thus allowing for high instructional density. The most accomplished teachers managed, on average, to engage virtually all of their students (96%) in the work of the classroom. By contrast, the moderately accomplished teachers achieved on-task rates of 84%, and the least accomplished teachers achieved on-task rates of 61%. This finding for our most accomplished teachers did not extend to the teachers in the most effective schools, who, as a group, did not differ from the teachers in the other schools in this study in terms of engagement rating.

Small-group instruction

Our finding that time spent in small-group instruction characterized the most accomplished teachers and the teachers in our most effective schools hearkens back to the important findings of the 1970s (Anderson et al., 1979; Stallings & Kaskowitz, 1974). The students in the most effective schools averaged 60 minutes a day of small, ability-grouped instruction. Here is a prime example of how classroom-level and school-level variables interact to produce a desirable outcome. The greater time allotted for small-group instruction did not just happen. It was made possible by the collaborative model used in all four schools, in which the classroom teacher, a resource teacher, an ESL teacher, and/or a special education teacher came together simultaneously and enabled every child, most typically, to have two blocks of small-group instruction.

Although ability-grouped instruction has been criticized in the past because it has been found to doom children to a lifetime in these low groups (Anderson, Hiebert, Scott, & Wilkinson, 1985; Barr & Dreeban, 1991; Gamoran, 1992; Hiebert, 1983; Oakes, 1985) and a persistently unambitious curriculum (Allington, 1983; Hiebert, 1983), it is important to remind ourselves of how teachers implemented grouping in these effective schools. Without

question, it was ability grouping. However, the use of systematic assessment prevented the groups from being rigid and inflexible. Besides establishing another layer of cultural barriers among students, the other major detriment of grouping is the differential nature of instruction accorded to different ability groups (see Allington, 1983). While differential treatment of different ability groups was not an explicit focus in our observations and teacher logs, there are data that indicate this did not occur in the four most effective schools. Finally, it is important to note that this is another example of the interaction of school-level factors (a common assessment system that enabled flexible movement between groups and a collaborative model that enabled flexible deployment of teaching personnel) and a classroom-level factor (small-group emphasis).

Independent reading

In the present study, we did find that the students in grades 1–3 in the most effective and moderately effective schools spent more time in independent reading (27–28 minutes per day) than the students in the least effective schools (19 mpd). These results support findings from earlier research that time spent in independent reading in school does make a difference in students' reading achievement (Anderson et al., 1988; Elley & Mangubhi, 1983; Taylor, Frye, & Maruyama, 1990). It is interesting to note that accomplished teachers did not allocate any more time to independent reading than their peers, suggesting that this is an instructional practice amenable to school-level influences.

Coaching

Although different terms have been used (e.g., use of structuring comments, probing of incorrect responses, scaffolded instruction), others have found this type of "on the fly" instruction to be a characteristic of effective teachers (Anderson et al., 1979; Brophy, 1973; Rosenshine & Furst, 1973; Wharton-MacDonald et al., 1998). Coaching also proved an important characteristic of the most accomplished teachers in the present study. More of the most accomplished teachers preferred coaching. While a coaching preference did not emerge as a general difference among teachers across school effectiveness ratings, we did find that the practice of coaching during reading to provide word recognition instruction was found to be a characteristic of teachers in the most effective schools and the most accomplished teachers in general. Perhaps coaching for word recognition instruction during children's reading of text is a place for teachers to begin to develop the ability to coach.

Phonics, yes, but phonics and...

The importance of systematic phonics instruction in learning to read has been repeatedly documented (Adams, 1990; Bond & Dykstra, 1967; Chall, 1967; Snow et al., 1998), as has the need for phonics to be developed in conjunction with real reading and writing (Adams, 1990). Wharton-MacDonald et al. (1998) found that the most effective first-grade teachers in their study taught decoding skills explicitly and provided their students with many opportunities to engage in authentic reading. However, systematic phonic instruction in isolation only, along with sheer opportunity to practice through reading connected text, may not be the optimal path toward a rich repertoire of word recognition strategies. Our data suggest that it is what teachers do to promote application of phonics knowledge during the reading of connected text that matters most. Our findings document that a majority of teachers in grades 1 and 2 across all schools taught phonics explicitly, in isolation. What distinguished the most accomplished teachers

and the majority of teachers in the most effective schools from their peers was their use of coaching to help students learn how to apply word recognition strategies to real reading. In kindergarten, the most accomplished teachers and teachers in the most effective schools were helping children apply their emerging phonemic awareness and phonics knowledge to the tracking and reading of big books and to writing. While more research is needed to unpack the specifics of these coaching and application strategies, our results suggest that conversations about systematic phonics instruction and opportunity to practice need to be broadened to include on the job coaching during everyday reading.

Higher level questions

Over twenty-five years ago, Rosenshine and Furst (1973) found that asking multiple levels of questions was consistently related to student achievement, and Puma et al. (1997) found that teachers in effective high-poverty schools emphasized both basic skills and higher order comprehension skills in reading. In the present study, we found that more of the most accomplished teachers frequently encouraged higher level responses to text. Strategies included asking higher level, aesthetic response questions and requiring students to write in response to what they had read. In comparison to teachers in the moderately and least effective schools, more teachers in the most effective schools focused on higher level questions. These findings lend support to this earlier body of research (Knapp, 1995) documenting the benefit of combining higher level and more basic strategy instruction. These encouraging differences notwithstanding, we must also remind ourselves that only 16% of the teachers in the entire sample could be considered to truly emphasize comprehension.

Improving instruction

In two of the most effective schools, teachers and principals mentioned the importance of improving instruction. Clearly, a number of the most effective schools had made an effort to improve classroom instruction, primarily through the intervention training they had put in place in their buildings. "Our early intervention program forced us to be of one accord. It has given teachers knowledge, strategies," reported the principal at Wheeler. When asked about advice to other schools, a teacher at Hilltop offered, "You need to focus staff development efforts on becoming better teachers of reading."

High expectations for student learning

In two of the most effective schools, teachers mentioned high expectations for students' achievement as a factor contributing to their success. Interestingly, one of these schools had the highest percentage of students on subsidized lunch (92%) in the entire sample. "I have an overarching belief that kids can learn and do learn to read by reading and writing. I have high expectations for them and help them create high expectations for themselves," reflected a Hilltop teacher. When asked what advice to give other schools, a Wheeler teacher exclaimed, "Start out with and keep the expectation that they can do it. We hear in the inner city they can't do it, but they CAN!"

Reading as a priority

When one looks across time spent in a variety of categories that fall roughly under the general rubric of reading instruction—whole-group instruction, small-group instruction, independent (seatwork) activities, independent reading, and writing in response to reading—the averages across levels of school effectiveness were: most effective—134 minutes; moderately effective—113 minutes; and least effective—113 minutes. Across all four most

effective schools, teachers were averaging 135 minutes a day on reading activities. Eighty-five minutes of this was either small-group or whole-group instruction, and almost 30 minutes of the total was independent reading. These times, based on teachers' logs from two different weeks, do not include time spent reading aloud to children, time spent in composition (in contrast to writing in response to reading), and time spent in spelling. The amount of time devoted to reading activities indicates that reading was an "operational" priority in the classrooms of the teachers in the most effective schools. In the words of one teacher at Stevenson, "My advice to other schools is let kids READ, READ, READ! WRITE, WRITE, WRITE! THINK, THINK, THINK!"

The Interaction of School and Classroom Factors

Interdependency and mutual support among people and practices

We uncovered many examples of the ways in which practices and resources interact to support quality instruction. But one particular cluster of practices surrounding the use of small groups is particularly provocative. Earlier, we argued that the presence of a schoolwide assessment system in these schools permitted them to implement small ability groups with sufficiently permeable boundaries to permit frequent between-group movement; we also argued that without a collaborative model in place, these schools would not have been able to deploy teacher personnel to find so much time for each small group. But the interaction does not end there. Students who spend more time in small groups are more likely to receive the coaching and scaffolding characteristic of effective classrooms and accomplished teachers. And this guidance and assisted performance, no doubt, contributed to the higher incidence of on-task behavior; students who know what is expected of them and receive help when they need it are more likely to be engaged. And from engagement, it is not hard to link to learning and achievement.

Teacher accomplishment and the possibilities for mentoring

While the argument is admittedly speculative, we want to suggest that the consistencies between our findings for teachers in the most effective schools and those for the most accomplished teachers may provide encouraging news for those who regard professional development as the center of gravity in any reform movement. Even though many of the practices of the most accomplished teachers in this study, such as coaching in word recognition during actual reading and asking higher level, aesthetic response questions, were mirrored in our analyses of teachers in the most effective schools, this does not mean that all of the most accomplished teachers worked in the most effective schools. In fact, only 58% of the teachers in the most effective schools were perceived from the observations to be most accomplished teachers (as compared to 30% in the least effective schools and 37% in the moderately effective schools). It is plausible, however, that these teachers were serving as models or coaches who brought particular areas of expertise to interactions with their colleagues. Our interviews certainly provided rich examples of this possibility. As the principal at Hilltop School explained, she had "worked to help people begin to appreciate the experts emerging within the building by bringing staff to the point where they acknowledged their expertise and by bringing teachers together to share their expertise and learn together."

Whatever the relationships among teachers (and we desperately need to learn more about how these relationships play themselves out and how to help skeptical teachers accept the belief that even the poorest children in their classes can learn), the fact that not every teacher in the most effective buildings is classified as a most accomplished teacher should be heartening to reformers who want to increase learning and achievement in our poorest schools. What it suggests is that a critical mass of highly accomplished teachers, which by our definition means teachers who possess more of the attributes of the canonical profile of culturally sensitive and pedagogically effective teachers, may be sufficient to move a school from the aspiring into the effective camp. Large-scale staff turnovers may not be necessary. There are exceptional teachers in all buildings who could be called upon to serve as models, peer coaches, or demonstration teachers to help committed teachers gain a concrete image of what positive instruction looks like and, in the process, improve their teaching.

Connections to the Effective Schools and Effective Teaching Traditions

That so many of our findings hearken back to earlier studies conducted in at least three different decades suggests, at first glance, that there may be little new information or insight in the current study. A second look, however, suggests that the current study adds much to this important and re-emerging line of inquiry.

First, in contrast to earlier studies, our work is more centrally grounded in reading curriculum and instructional practices than the bulk of the effective schools and effective teaching work. Our measures, our process for taking field notes, our surveys, our interview protocols, and our observational tools are all solidly grounded in reading curriculum and instructional matters rather than more general conceptualizations of teaching and program organization. Thus our findings have a much more specific cast and character to them. For example, we were able to add detail, relevant to reading programs and student reading performance, to earlier findings on systematic evaluation, home-school connections, building collaboration, student engagement, and teacher scaffolding.

Second, because we were able to combine both school-level and classroom-level analyses of programs and practices, we learned more about how these two levels of analysis and implementation support or interfere with one another than has been possible in studies that focus on one or the other. As a prime example, we would mention the less-than-perfect correlation between effective schools and accomplished teachers: not all of the latter are in the former. As a second example, we would mention the fact that the capacity to engage students in learning seems more a teacher than a program characteristic. The interaction between collaboration and assessment, both school-level factors, and small-group emphasis, a classroom-level factor, provides a third, and perhaps the most compelling example of this interaction between building and classroom factors.

Third, we have benefited from a simple accident of history. In 1999, because we have built our research capacity by standing on the shoulders of our predecessors, we have more and better tools available to tease out these puz-

zling relationships. The fact, for example, that we were able to use a wide range of complementary data sources (student performance, observation protocols, field notes, interviews, and surveys) helped us to pin down and particularize the relationships between program and teaching characteristics and student performance.

Limitations and Conclusions

While we are encouraged by the current findings, we must remind ourselves that this work, like all of the work in the effective schools and effective teaching tradition, comes with serious limitations. When all is said and done, we are examining natural correlations that exist between program and teaching factors on the one hand and student performance on the other. While these correlations may be useful in planning more definitive research and in guiding the development of local programs and policies, they cannot be used to fix causes for improvements (or decrements) in student achievement. For that, we need more systematic experimentation, including control groups, randomization, and careful analyses of growth over time. It is to that agenda that we will soon turn our attention.

Our work carries a number of additional, more specific limitations. We would have liked to assess more students per classroom simply to improve the precision and trustworthiness of our work. Because we did not test the full range of students within classrooms, we were unable to examine aptitude by treatment interactions.

Also, our prior information about schools was unintentionally misleading. In terms of selection, schools that we had expected to rise to the top of our achievement scales (because of their reputations) did not always do so. Conversely, some schools that were thought to be ordinary in terms of achievement and reform did better than our information would have led us to believe. What this suggests is that the static assumptions about school status that we used to select schools were inappropriate. All of these schools were, and are, on the move, in one direction or another. Curriculum leaders and teaching staffs come and go, and with them the energy to initiate or sustain schoolwide reform.

Our measures were not perfect. We would have liked to use more and better measures of a wider range of skills and strategies. We would have liked a more useful writing measure, and, in our future work, we will ensure that we get reliable information on this important aspect of literacy acquisition. Furthermore, those who adopt a psychometric lens would take issue with our reliance on classroom-based assessments administered by multiple assessors in multiple sites. Most informal inventories, writing samples, word lists, and even the available tests of phonemic awareness do not possess the psychometric underbelly of standardized multiple-choice tests. Those who would adopt a lens of authenticity would be equally disappointed in our measures. While they are classroom-based, they are not the stuff of constructivist-based reform. Lacking is any appreciation of response to literature and personal engagement with text. Those who adopt a cultural lens would find both our student and our teacher measures wanting. They would find that our student measures are not likely to be sensitive to the special

skills or perspectives that children develop in culturally rich settings. And they would find that our observational lenses did not guarantee that observers would look directly for culturally responsive (or culturally insensitive) instruction. In defense of what we did, all we can say is that, like all school-based research efforts, we made compromises motivated by cost and credibility. Because we were personally and painfully aware of the problems of using group assessments, especially with kindergartners and first graders, and because we wanted assessments that would be credible with the teachers in these schools, we were committed to one-on-one assessments, oral reading samples, and retellings—all practices we knew would appeal to teachers.

Beating the odds predicted by poverty takes dedication and hard work. A school needs good morale and teachers willing to work tirelessly and collaboratively. Teachers and schools also have to go the extra mile to reach out to parents. The results of this study suggest that children in the primary grades make the greatest growth when a high proportion of their reading instruction is delivered through small ability groups, when their progress is monitored regularly, and when they have ample time to read and to learn needed skills and strategies. Teachers who are most accomplished in helping children thrive in reading are skilled in coaching and in keeping all children academically engaged. The findings of this study also suggest that schools, even our most effective schools, have a long way to go in improving reading instruction in the primary grades.

It is clear from this study that a combination of sound building decisions and collaborative efforts as well as effective practices within individual classrooms are needed if schools are to succeed at beating the odds in terms of primary grade students' reading achievement. From the descriptive analyses and case studies, it is clear that the process of becoming a beat the odds school is complex; equally complex are the conditions that characterize a beat the odds school. Staff in all of these exceptional schools talked about the work they had been doing over a number of years to improve and the work that still lay ahead of them. In other words, there is no single quick answer to the question of how best to reshape a school's reading program and the repertoire of instructional practices teachers employ in of the quest of helping all children read well by the time they leave the primary grades. What is equally clear is the educators who combine high expectations for all their children with a thirst for improving their classroom practice, a commitment to strong, collaboratively forged school-wide programs, and plain old-fashioned hard work can meet great expectations for the children with and for whom they work.

In all of the most effective schools and most of the other schools in this study, the building environments were positive, and the schools were friendly places for children to learn. In most schools, the teachers and principals were genuinely concerned about developing the abilities of struggling readers, and in many, appeared willing to take necessary steps to improve the reading achievement of all their students. The question that lies before us, and one to which we are currently turning our attention, is how (and whether) we, as a profession, can import the sorts of values and practices found in the most effective schools to other schools that only aspire to these levels of accomplishment.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Allington, R. L. (1983). The reading instruction provided readers of differing abilities. *Elementary School Journal, 83,* 548–559.

American Federation of Teachers. (1997). *Raising student achievement: A resource guide for redesigning low-performing schools.* (AFT Item Number 3780). Washington DC: Author.

Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79,* 193–223.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers.* Washington, DC: National Institute of Education.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23* (3), 285–303.

Barr, R., & Dreeban, R. (1991). Grouping students for reading instruction. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, Vol. II* (pp. 885–910). New York: Longman.

Bond, G. J. & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2,* 5–142.

Brophy, J. (1973). Stability of teacher effectiveness. *American Educational Research Journal, 10,* 245–252.

Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillip, G. W. (1996). *NAEP 1994 reading report card for the nation and the states.* Washington, DC: Office of Educational Research and Improvement.

Carroll, J. B., Davies, P. & Richman, B. (1971). *The American heritage word frequency book.* Boston: Houghton Mifflin.

Chall, J. S. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Colt, J. (1997). *A scoring rubric for children's story retelling.* Longmont, CO: St. Vrain Valley School District.

Cunningham, P. & Cunningham, J. (1992). Making words: Enhancing the invented spelling-decoding connection. *Reading Teacher, 46,* 106–107.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232.

Donahue, P. L., Voelkl, K. E., Campbell, J. R., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation.* Washington, DC: U.S. Department of Education.

Elley, W. B., & Mangubhi, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly, 19*, 53-67.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham and A. Lieberman (Eds.), *Time to learn*. Washington, DC: National Institute of Education.

Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership, 50* (2), 11-17.

Herman, R, (1999). *An educator's guide to schoolwide reform*. Washington, DC: American Institutes for Research.

Herman, R., & Stringfield, S. (1997). *Ten promising programs for educating all children: Evidence of impact*. Arlington, VA: Educational Research Service.

Hiebert, E. H. (1983). An examination of ability grouping for reading instruction. *Reading Research Quarterly, 8*, 231-255.

Hiebert, E. H., Colt, J. M., Catto, S. L., & Gury, E. C. (1992). Reading and writing of first-grade students in a restructured Chapter 1 program. *American Educational Research Journal, 29*, 545-572.

Hirsch, E. D. (1987). *Cultural literacy: What every American needs to know*. New York: Vintage.

Hoffman, J. V. (1991). Teacher and school effects in learning to read. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, Vol. II* (pp. 911-950). New York: Longman.

Invernizzi, M., Juel, C., & Rosemary, C. A. (1997). A community volunteer tutorial that works. *Reading Teacher, 50* (4), 304-311.

Knapp, M. S. (Ed.). (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.

Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American children*. San Francisco: Josey-Bass.

Leslie, L., & Caldwell, J. (1995). *Qualitative reading inventory—II*. New York: Harper Collins.

Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30*, 123-148.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Pikulski, J. (1996) *Emergent literacy survey*. Boston: Houghton Mifflin.

Pinnell, G. S., Lyons, C.A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29*, 8-39.

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary grade teachers nominated as effective in promoting literacy. *Elementary School Journal, 96,* 363-384.

Puma, M. J., Karweit, N., Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes.* Washington, DC: Planning and Evaluating Service, U.S. Department of Education.

Roehler, L., & Duffy, G. (1991). Teachers' instructional actions in learning to read. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, Vol. II* (pp. 861-884). New York: Longman.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching* (pp. 122-183). Chicago: Rand McNally.

Shavelson, R. J., & Berliner, D. C. (1988). Evasion of the education research infrastructure: A reply to Finn. *Educational Researcher, 17,* 9-11.

Slavin, R. E., & Fashola, O. S. (1998). *Show me the evidence! Proven and promising programs for America's schools.* Thousand Oaks, CA: Corwin Press.

Snow, C. E., Burns, S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children: Report of the Committee on the Prevention of Reading Difficulties in Young Children.* Washington, DC: National Academy Press.

Stallings, J., & Kaskowitz, D. (1974). *Follow through classroom observation evaluation 1972-1973* (SRI Project URU-7370). Stanford, CA: Stanford Research Institute.

Stringfield, S., Millsap, M. A., & Herman, R. (1997). *Urban and suburban/rural special strategies for educating disadvantaged children: Findings and policy implications of a longitudinal study.* Washington, DC: U.S. Department of Education.

Taylor, B. M., Frye, B. J., & Maruyama, G. M. (1990). Time spent in reading and reading growth. *American Educational Research Journal, 27* (2), 351-362.

Taylor, B. M., Short, R. A., Frye, B. J., & Shearer, B. (1992). Classroom teachers prevent reading failure among low-achieving first grade students. *Reading Teacher, 45,* 592-597.

Viisc, N. (1994). *Word feature spelling list.* Charlottesville: University of Virginia Bookstore.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1998). *What do we know: Widely implemented school improvement programs.* Philadelphia, PA: Center for Research in Human Development and Education.

Weber, G. (1971). *Inner city children can be taught to read: Four successful schools* (CGE Occasional Papers No. 18; ERIC Document Reproduc-

tion Service No. Ed 057 125). Washington, DC: Council for Basic Education.

Wharton-MacDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal, 99,* 101-128.

Wilder, G. (1977). Five exemplary reading programs. In J. T. Guthrie (Ed.), *Cognition, curriculum, and comprehension* (pp. 57-68). Newark, DE: International Reading Association.

Appendix A

QRI Retelling Sheet

Passage _____

Child _____

Retelling: Ask, "What happened in this story?"

Record student's response as best as possible:

Score:

1	2	3	4
• Student offers little or no information about the selection. • Retelling is incomprehensible. • Stated ideas do not relate to the selection.	• Student relates details only. • Student is unable to recall the gist of the selection. • Retelling is incomplete or ideas are misconstrued. • Sequence is not logical.	• Student relates some main ideas and some supporting details. • Retelling is fairly coherent. • Sequence is logical.	• All major points and appropriate supporting details are included. • There is a high degree of completeness and coherence. • Student generalizes beyond the text.

Appendix B

Child Name _____ Grade___ Examiner ID ___ Date _____

and	___	there	___	each	___	still	___	complete	___
to	___	do	___	like	___	food	___	anything	___
you	___	how	___	through	___	room	___	wear	___
that	___	about	___	new	___	money	___	sheep	___
was	___	some	___	good	___	morning	___	nation	___
they	___	these	___	any	___	noticed	___	blow	___
his	___	would	___	right	___	begins	___	peace	___
at	___	has	___	also	___	weather	___	climate	___
from	___	him	___	come	___	friend	___	rough	___
I	___	see	___	because	___	sent	___	struck	___
not	___	could	___	does	___	insects	___	speaking	___
had	___	make	___	say	___	trade	___	magic	___
what	___	who	___	give	___	clock	___	lion	___
all	___	get	___	air	___	gate	___	crowded	___
an	___	look	___	boy	___	pain	___	removed	___
said	___	big	___	mother	___	breathe	___	wool	___
man	___	home	___	point	___	pride	___	worried	___
stop	___	red	___	move	___	promise	___	claws	___
map	___	run	___	true	___	clue	___	stamps	___
bad	___	dog	___	road	___	hatch	___	senses	___

Preprimer		Primer		First		Second		Third	
Total	___	Total	___	Total	___	Total	___	Total	___

Instructions for Administering Word Lists

Examiner instructions

Directions for administering word lists are the same for all grade levels. Begin administration with the Preprimer list and have the child continue reading from one list to the next until he or she makes seven (7) consecutive errors. Count the number of correct responses for each level and record the number at the bottom of each sheet.

Continue to administer a new list in this fashion until the child misses 7 or more words. Give the child time to attempt each word. Then uncover the remaining lists, and put a check next to any additional word that the child reads. Count those words and record the totals at the bottom of the record sheets.

Instructions for the children

I have some lists of words that I want you to read one at a time. Some of the words will be easy for you and some I expect to be very hard. Don't worry. I don't expect you to know all of them. I cannot help you because I want to see what you can do on your own. Do your very best. Look right above this card. That's where you will see each word. Are you ready?

(and after the child has missed 7 consecutive words on any list)
Thanks for your reading. Now I'm going to let you look at the rest of my words. Are there any that you know?

Appendix C

The reader is referred to authors, in press, for a full account of the kindergarten analysis. In brief, these are the highlights of the findings. Students of the most accomplished kindergarten teachers had significantly higher spring phonemic awareness scores than students of the least accomplished teachers. Analysis of logs revealed that the most accomplished teachers (when considering all-day and half-day programs separately) spent more time on phonemic awareness and phonics activities than the least accomplished teachers. More of the most accomplished teachers were frequently observed practicing with their children how to track during reading of big books, stories, or charts than the least accomplished teachers. More of the teachers in the most effective schools were frequently observed having their children write for sounds during activities such as journal writing or guided writing than teachers in the least effective schools.

About CIERA

The Center for the Improvement of Early Reading Achievement (CIERA) is the national center for research on early reading and represents a consortium of educators in five universities (University of Michigan, University of Virginia, and Michigan State University with University of Southern California and University of Minnesota), teacher educators, teachers, publishers of texts, tests, and technology, professional organizations, and schools and school districts across the United States. CIERA is supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

Mission. CIERA's mission is to improve the reading achievement of America's children by generating and disseminating theoretical, empirical, and practical solutions to persistent problems in the learning and teaching of beginning reading.

CIERA Research Model

The model that underlies CIERA's efforts acknowledges many influences on children's reading acquisition. The multiple influences on children's early reading acquisition can be represented in three successive layers, each yielding an area of inquiry of the CIERA scope of work. These three areas of inquiry each present a set of persistent problems in the learning and teaching of beginning reading:

CIERA INQUIRY 1
Readers and Texts

Characteristics of readers and texts and their relationship to early reading achievement. What are the characteristics of readers and texts that have the greatest influence on early success in reading? How can children's existing knowledge and classroom environments enhance the factors that make for success?

CIERA INQUIRY 2
Home and School

Home and school effects on early reading achievment. How do the contexts of homes, communities, classrooms, and schools support high levels of reading achievement among primary-level children? How can these contexts be enhanced to ensure high levels of reading achievement for all children?

CIERA INQUIRY 3
Policy and Profession

Policy and professional effects on early reading achievement. How can new teachers be initiated into the profession and experienced teachers be provided with the knowledge and dispositions to teach young children to read well? How do policies at all levels support or detract from providing all children with access to high levels of reading instruction?

www.ciera.org

Center for the Improvement of
Early Reading Achievement

CIERA is a collaboration of
University of Michigan
University of Virginia
Michigan State University

with
University of Minnesota
University of Southern California

University of Michigan School of Education
610 E University Av, Rm 1600 SEB
Ann Arbor, MI 48109-1259

734.647.6940 voice
734.763.1229 fax
ciera@umich.edu

www.ciera.org